

Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/US2004/029541

International filing date: 09 September 2004 (09.09.2004)

Document type: Certified copy of priority document

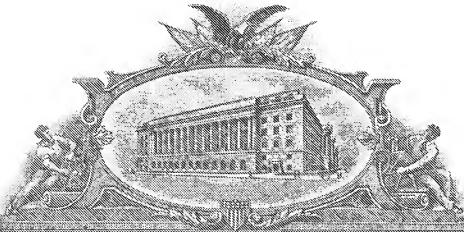
Document details: Country/Office: US
Number: 60/501,788
Filing date: 10 September 2003 (10.09.2003)

Date of receipt at the International Bureau: 14 March 2007 (14.03.2007)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)



World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse



THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

March 05, 2007

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: 60/501,788

FILING DATE: *September 10, 2003*

RELATED PCT APPLICATION NUMBER: *PCT/US04/29541*

THE COUNTRY CODE AND NUMBER OF YOUR PRIORITY APPLICATION, TO BE USED FOR FILING ABROAD UNDER THE PARIS CONVENTION, IS *US60/501,788*



Certified by

Under Secretary of Commerce
for Intellectual Property
and Director of the United States
Patent and Trademark Office

This is a request for filing a **PROVISIONAL APPLICATION FOR PATENT** under 37 CFR 1.53(c).

Express Mail Label No. **EV308275477**

INVENTOR(S)					
Given Name (first and middle [if any])		Family Name or Surname		Residence (City and either State or Foreign Country)	
Gary M.		Fader		Landenberg, PA - U.S.	
Additional Inventors are being named on the <u>1</u> separately numbered sheets attached hereto					
TITLE OF THE INVENTION (500 characters max)					
DIHYDROFLAVONOL-4-REDUCTASE					
Direct all correspondence to: CORRESPONDENCE ADDRESS					
<input checked="" type="checkbox"/> Customer Number: 23906					
OR					
<input type="checkbox"/> Firm or Individual Name					
Address					
Address					
City		State		Zip	
Country		Telephone		Fax	
ENCLOSED APPLICATION PARTS (check all that apply)					
<input checked="" type="checkbox"/> Specification Number of Pages <u>38</u>					
<input checked="" type="checkbox"/> Drawing(s) Number of Sheets <u>7</u>					
<input type="checkbox"/> Application Date Sheet. See 37 CFR 1.76					
<input type="checkbox"/> CD(s), Number _____					
<input checked="" type="checkbox"/> Other (specify) Seq. Listing-CRF and paper copy-20 pgs., Seq. Statement					
METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT					
<input type="checkbox"/> Applicant claims small entity status. See 37 CFR 1.27.					
<input type="checkbox"/> A check or money order is enclosed to cover the filing fees.					
<input checked="" type="checkbox"/> The Director is hereby authorized to charge filing fees or credit any overpayment to Deposit Account Number: <u>04-1928</u>					
<input type="checkbox"/> Payment by credit card. Form PTO-2038 is attached.					
FILING FEE Amount (\$) \$160.00					
The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.					
<input checked="" type="checkbox"/> No.					
<input type="checkbox"/> Yes, the name of the U.S. Government agency and the Government contract number are: _____					

[Page 1 of 2]

Respectfully submitted,

SIGNATURE

TYPED or PRINTED NAME J KENNETH JOUNG

TELEPHONE (302) 992-4929

Date 10 September 2003

REGISTRATION NO. 41,881

(if appropriate) Docket Number: BB1521 US PRV

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

This collection of information is required by 37 CFR 1.51. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 8 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Mail Stop Provisional Application, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

60/501788
 09/10/03

PROVISIONAL APPLICATION COVER SHEET
Additional Page

PTO/SB/16 (08-03)

Approved for use through 07/31/2008, OMB 0651-0032

U.S. Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

Docket Number **BB1521 US PRV**

INVENTOR(S)/APPLICANT(S)		
Given Name (first and middle (if any))	Family or Surname	Residence (City and either State or Foreign Country)
Omolayo O.	Famodu	Newark, DE - U.S.
Lisa L.	Huang	Hockessin, DE- Peoples Republic of China
Brian	McGonigle	Wilmington, DE - U.S.
Yong	Tao	Newark, DE - Peoples Republic of China

[Page 2 of 2]

Number _____ of _____

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

16310 U.S. PTO
09/10/03

Approved for use through 07/31/2006. OMB 0651-0032
U.S. Patent and Trademark Office, U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

FEE TRANSMITTAL for FY 2003

Effective 01/01/2003. Patent fees are subject to annual revision.

☐ Applicant claims small entity status. See 37 CFR 1.27

TOTAL AMOUNT OF PAYMENT (\$) 160.00

Complete if Known

Application Number	To be Assigned
Filing Date	Concurrently Herewith
First Named Inventor	Gary M. Fader et al.
Examiner Name	To be Assigned
Art Unit	Unknown
Attorney Docket No.	BB1521 USPRV

METHOD OF PAYMENT (check all that apply)

☐ Check ☐ Credit card ☐ Money Order ☐ Other ☐ None

☒ Deposit Account:

Deposit Account Number: 04-1928
Deposit Account Name: E. I. du Pont de Nemours and Company

The Director is authorized to: (check all that apply)

☒ Charge fee(s) indicated below ☒ Credit any overpayments
☒ Charge any additional fee(s) during the pendency of this application
☒ Charge fee(s) indicated below, except for the filing fee to the above-identified deposit account.

FEE CALCULATION

1. BASIC FILING FEE

Large Entity Fee Code (\$)	Small Entity Fee Code (\$)	Fee Description	Fee Paid
1001 750	2001 375	Utility filing fee	
1002 330	2002 165	Design filing fee	
1003 520	2003 260	Plant filing fee	
1004 750	2004 375	Reissue filing fee	
1005 160	2005 80	Provisional filing fee	160.00
SUBTOTAL (1)			160.00

2. EXTRA CLAIM FEES FOR UTILITY AND REISSUE

	Extra Claims	Fee from below	Fee Paid
Total Claims	-20+ =	X 18	
Independent Claims	-3+ =	X 84	
Multiple Dependent	YES	280	

Large Entity - Small Entity

Fee Code (\$)	Fee Code (\$)	Fee Description
1202 18	2202 9	Claims in excess of 20
1201 84	2201 42	Independent claims in excess of 3
1203 280	2203 140	Multiple dependent claim, if not paid
1204 84	2204 42	** Reissue independent claims over original patent
1205 18	2205 9	** Reissue claims in excess of 20 and over original patent

SUBTOTAL (2) (\$) 0.00

**or number previously paid, if greater. For Reissues, see above

FEE CALCULATION (continued)

3. ADDITIONAL FEES

Large Entity - Small Entity

Fee Code (\$)	Fee Code (\$)	Fee Code (\$)	Fee Description	Fee Paid
1051 130	2051 85		Surcharge - late filing fee or oath	
1052 50	2052 25		Surcharge - late provisional filing fee or cover sheet	
1053 130	1053 130		Non-English specification	
1812 2,520	1812 2,520		For filing a request for <i>ex parte</i> reexamination	
1804 920*	1804 920*		Requesting publication of SIR prior to Examiner action	
1805 1,840*	1805 1,840*		Requesting publication of SIR after Examiner action	
1251 110	2251 55		Extension for reply within first month	
1252 410	2252 205		Extension for reply within second month	
1253 930	2253 465		Extension for reply within third month	
1254 1,450	2254 725		Extension for reply within fourth month	
1255 1,970	2255 985		Extension for reply within fifth month	
1401 320	2401 160		Notice of Appeal	
1402 320	2402 160		Filing a brief in support of an appeal	
1403 280	2403 140		Request for oral hearing	
1451 1,510	1451 1,510		Petition to institute a public use proceeding	
1452 110	2452 55		Petition to revive - unavoidable	
1453 1,300	2453 650		Petition to revive - unintentional	
1501 1,300	2501 650		Utility issue fee (or reissue)	
1502 470	2502 235		Design issue fee	
1503 630	2503 315		Plant issue fee	
1460 130	1460 130		Petitions to the Commissioner	
1807 50	1807 50		Processing fee under 37 CFR 1.17(q)	
1806 180	1806 180		Submission of Information Disclosure Stmt	
8021 40	8021 40		Recording each patent assignment per property (times number of properties)	
1809 750	2809 375		Filing a submission after final rejection (37 CFR 1.129(a))	
1810 750	2810 375		For each additional invention to be examined (37 CFR 1.129(b))	
1801 750	2801 375		Request for Continued Examination (RCE)	
1802 900	1802 900		Request for expedited examination of a design application	

Other fee (specify)

*Reduced by Basic Filing Fee Paid

SUBTOTAL (3) (\$) 0.00

SUBMITTED BY

Name (Print/Type) J Kenneth Joung

Signature

Registration No.

41,881

(Complete if applicable)

Telephone

(302) 992-4929

Date

10 Sept 2003

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

This collection of information is required by 37 CFR 1.17 and 1.27. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

TITLE
 DIHYDROFLAVONOL-4-REDUCTASE
FIELD OF THE INVENTION

The field of invention relates to plant molecular biology, and more specifically, to nucleic acid fragments encoding dihydroflavonol-4-reductase in plants and seeds.

BACKGROUND OF THE INVENTION

The phenylpropanoid pathway provides plants with a mechanism for the conversion of large amounts of carbon from aromatic amino acid metabolism into the biosynthesis of products based on a phenylpropane skeleton. Progress has been made in gaining an understanding of phenylpropanoid metabolism by manipulating some of the components of its pathway in transgenic plants. The products of phenylpropanoid metabolism provide highly efficient UV screening (flavonoids and small phenolic compounds), control of water loss and structural rigidity (lignin), and defense against pathogens (many phenylpropanoid compounds), all traits that are essential for the success of land plants.

An important branch of phenylpropanoid metabolism is the flavonoid biosynthesis pathway. Flavonoids represent one of the major secondary metabolite families in plants and comprise several subclasses. The biosynthetic pathway that leads to the production of flavonoids is highly branched and complex. The first committed step in the pathway involves the condensation of three molecules of malonyl-CoA with one molecule of *p*-coumaroyl-CoA (Figures 1A and 1B). This reaction is catalyzed by the enzyme chalcone synthase. The product of this reaction, 2',4,4',6'-tetrahydroxychalcone, is normally rapidly isomerized to produce naringenin by the enzyme chalcone isomerase (also known as chalcone-flavanone isomerase). Naringenin is subsequently hydroxylated at the 3-position of the central ring by flavanone-3-hydroxylase to produce dihydrokaempferol. The B-ring of dihydrokaempferol can be hydroxylated at either the 3', or both the 3' and 5' positions, to produce dihydroquercetin and dihydromyricetin, respectively.

The next key enzyme in one of the pathways is dihydroflavonol-4-reductase (E.C. 1.1.1.219; also known as dihydrokaempferol-4-reductase). This enzyme converts dihydroflavonols to leucoanthocyanidins (see, e.g., Figure 1B). Dihydroflavonol-4-reductase has variable substrate specificity depending on its plant source and has the potential to act on any one or more of dihydrokaempferol, dihydroquercetin and dihydromyricetin. To date, a few dihydroflavonol-4-reductase genes have been isolated from plants, including rice, alfalfa, maize, petunia and strawberry (Zhuang et al., *Plant Physiol.* 120:633 (1999) - PGR99-074; Beld et al., *Plant Mol. Biol.* 13:491-502 (1989); Bernhardt et al., *Plant J.* 14:483-488 (1998);

Charrier et al., *Plant Mol. Biol.* 29:773-786 (1995); Moyano et al., *Plant Physiol.* 117:711-716 (1998)). The present invention provides novel dihydroflavonol-4-reductases from soybean (*Glycine max*) and wheat (*Triticum aestivum*).

SUMMARY OF THE INVENTION

The present invention relates to isolated polynucleotides comprising a nucleotide sequence encoding a polypeptide having dihydroflavonol-4-reductase activity, wherein the amino acid sequence of the polypeptide and the amino acid sequence of SEQ ID NO:4, 6, 8, 12, 14, or 16 preferably have at least 80%, 85%, 90%, 95%, or 96% sequence identity. The present invention also relates to isolated polynucleotides comprising the complement of the nucleotide sequence. The present invention also relates to isolated polynucleotides encoding the polypeptide sequence of SEQ ID NO:4, 6, 8, 12, 14, or 16 or nucleotide sequences comprising the nucleotide sequence of SEQ ID NO:3, 5, 7, 11, 13 or 15.

In a preferred first embodiment, an isolated polynucleotide comprises: (a) a nucleotide sequence encoding a polypeptide having dihydroflavonol-4-reductase activity, wherein the polypeptide has an amino acid sequence of at least 80%, 85%, 90%, 95%, or 96% sequence identity, based on the Clustal V method of alignment, when compared to one of SEQ ID NO:4, 6, 8, 12, 14, or 16, or (b) a complement of the nucleotide sequence, wherein the complement and the nucleotide sequence consist of the same number of nucleotides and are 100% complementary. The polypeptide preferably comprises the amino acid sequence of SEQ ID NO:4, 6, 8, 12, 14, or 16. The nucleotide sequence preferably comprises the nucleotide sequence of SEQ ID NO:3, 5, 7, 11, 13 or 15.

In a preferred second embodiment, a recombinant DNA construct comprises any of the isolated polynucleotides of the present invention operably linked to at least one regulatory sequence, and a cell, a plant, and a seed comprising the recombinant DNA construct.

In a preferred third embodiment, a vector comprises any of the isolated polynucleotides of the present invention.

In a preferred fourth embodiment, a method for transforming a cell comprises transforming a cell with any of the isolated polynucleotides of the present invention. The cell transformed by this method is also included. Advantageously, the cell is eukaryotic, e.g., a yeast or plant cell, or prokaryotic, e.g., a bacterium.

In a preferred fifth embodiment, a method for producing a transgenic plant comprises transforming a plant cell with any of the isolated polynucleotides or recombinant DNA constructs of the present invention and regenerating a plant from

the transformed plant cell. Also included is a transgenic plant produced by this method, and seed obtained from this transgenic plant.

In a preferred sixth embodiment, an isolated polypeptide having dihydroflavonol-4-reductase activity, wherein the polypeptide has an amino acid sequence of at least 80%, 85%, 90%, 95% or 96% identity, based on the Clustal V method of alignment, when compared to one of SEQ ID NO:4, 6, 8, 12, 14, or 16. The polypeptide preferably comprises one of SEQ ID NO:4, 6, 8, 12, 14, or 16.

In a preferred seventh embodiment, a method for isolating a polypeptide having dihydroflavonol-4-reductase activity comprises transforming a cell with a recombinant DNA construct of the present invention, growing the cell in a culture medium, and isolating the polypeptide from the cell or culture medium.

In a preferred eighth embodiment, a method for selecting a transformed cell comprises: (a) transforming a host cell with the recombinant DNA construct or an expression cassette of the present invention; and (b) growing the transformed host cell, preferably a plant cell, under conditions that allow expression of the dihydroflavonol-4-reductase polynucleotide in an amount sufficient to complement a null mutant in order to provide a positive selection means.

In a preferred ninth embodiment, a method of altering the level of expression of a dihydroflavonol-4-reductase protein in a host cell comprises: (a) transforming a host cell with a recombinant DNA construct of the present invention; and (b) growing the transformed host cell under conditions that are suitable for expression of the recombinant DNA construct wherein expression of the recombinant DNA construct results in production of altered levels of the dihydroflavonol-4-reductase protein in the transformed host cell.

BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE LISTINGS

The invention can be more fully understood from the following detailed description and the accompanying drawings and Sequence Listing which form a part of this application.

Figures 1A and 1B depict portions of the flavonoid biochemical pathway.

Figures 2A, 2B and 2C show a comparison of: (i) the amino acid sequences of the dihydroflavonol-4-reductase encoded by the following: (a) nucleotide sequence derived from soybean clone sre.pk0043.d11:fis (SEQ ID NO:4), (b) nucleotide sequence derived from soybean clone sls2c.pk027.c22:fis (SEQ ID NO:6), (c) nucleotide sequence derived from soybean clone ssl1c.pk004.m3:fis (SEQ ID NO:8); (ii) amino acid sequence of *Glycine max* dihydroflavonol-4-reductase (NCBI General Identifier No. 5852933; SEQ ID NO:17); and (iii) amino

acid sequence from *Arabidopsis thaliana* dihydroflavonol-4-reductase (NCBI General Identifier No. 6689826; SEQ ID NO:18). Dashes are used by the program to maximize alignment of the sequences.

Figures 3A and 3B show a comparison of: (i) the amino acid sequences of the dihydroflavonol-4-reductase encoded by the following: (a) nucleotide sequence derived from wheat clone wdk3c.pk012.j12:fis (SEQ ID NO:12), (b) nucleotide sequence derived from wheat clone wdi1c.pk006.d20:fis (SEQ ID NO:14), (c) nucleotide sequence derived from wheat clone wdk2c.pk017.h12:fis (SEQ ID NO:16); (ii) amino acid sequence of *Hordeum vulgare* dihydroflavonol-4-reductase (NCBI General Identifier No. 1706372; SEQ ID NO:19); and (iii) amino acid sequence from *Oryza sativa* dihydroflavonol-4-reductase (NCBI General Identifier No. 4581047; SEQ ID NO:20). Dashes are used by the program to maximize alignment of the sequences.

Table 1 lists the polypeptides that are described herein, the designation of the cDNA clones that comprise the nucleic acid fragments encoding polypeptides representing all or a substantial portion of these polypeptides, and the corresponding identifier (SEQ ID NO:) as used in the attached Sequence Listing. The sequence descriptions and Sequence Listing attached hereto comply with the rules governing nucleotide and/or amino acid sequence disclosures in patent applications as set forth in 37 C.F.R. §1.821-1.825.

TABLE 1
Dihydroflavonol-4-Reductase

Protein	Clone Designation	SEQ ID NO:	
		(Nucleotide)	(Amino Acid)
Soybean polypeptide similar to <i>Fragaria ananassa</i> dihydroflavonol-4-reductase	ssl1c.pk004.m3	1	2
Soybean polypeptide similar to <i>Glycine max</i> dihydroflavonol-4-reductase	sre.pk0043.d11:fis	3	4
Soybean polypeptide similar to <i>Glycine max</i> dihydroflavonol-4-reductase	sls2c.pk027.c22:fis	5	6
Soybean polypeptide similar to <i>Arabidopsis</i> dihydroflavonol-4-reductase	ssl1c.pk004.m3:fis	7	8

Wheat polypeptide similar to <i>Zea mays</i> dihydroflavonol-4-reductase	wdi1c.pk006.d20	9	10
Wheat polypeptide similar to <i>Hordeum vulgare</i> dihydroflavonol-4-reductase	wdk3c.pk012.j12:fis	11	12
Wheat polypeptide similar to <i>Oryza sativa</i> dihydroflavonol-4-reductase	wdi1c.pk006.d20:fis	13	14
Wheat polypeptide similar to <i>Hordeum vulgare</i> dihydroflavonol-4-reductase	wdk2c.pk017.h12:fis	15	16

SEQ ID NO:17 is the amino acid sequence of a dihydroflavonol-4-reductase from *Glycine max* (NCBI General Identifier No. 5852933).

SEQ ID NO:18 is the amino acid sequence of a dihydroflavonol-4-reductase from *Arabidopsis thaliana* (NCBI General Identifier No. 6689826).

SEQ ID NO:19 is the amino acid sequence of a dihydroflavonol-4-reductase from *Hordeum vulgare* (NCBI General Identifier No. 1706372).

SEQ ID NO:20 is the amino acid sequence of a dihydroflavonol-4-reductase from *Oryza sativa* (NCBI General Identifier No. 4581047).

SEQ ID NO:21 is the amino acid sequence of a dihydroflavonol-4-reductase from *Fragaria ananassa* (NCBI General Identifier No. 2599562).

SEQ ID NO:22 is the amino acid sequence of a dihydroflavonol-4-reductase from *Zea mays* (NCBI General Identifier No. 14030554).

The Sequence Listing contains the one letter code for nucleotide sequence characters and the three letter codes for amino acids as defined in conformity with the IUPAC-IUBMB standards described in *Nucleic Acids Res.* 13:3021-3030 (1985) and in the *Biochemical J.* 219 (No. 2):345-373 (1984) which are herein incorporated by reference. The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

DETAILED DESCRIPTION OF THE INVENTION

The disclosure of each patent, patent application and non-patent literature in the instant specification is incorporated herein by reference in its entirety.

In the context of this disclosure, a number of terms shall be utilized. The term "dihydroflavonol-4-reductase" refers to the polypeptide or enzyme that catalyzes the conversion of dihydroflavonols into leucoanthocyanidins.

Dihydroflavonol-4-reductase has variable substrate specificity depending on its plant source and has the potential to act on any one or more of dihydrokaempferol, dihydroquercetin and dihydromyricetin. Dihydroflavonol-4-reductase is also known as dihydrokaempferol-4-reductase and DFR. These terms can be used interchangeably.

The terms "polynucleotide", "polynucleotide sequence", "nucleic acid sequence", and "nucleic acid fragment"/"isolated nucleic acid fragment" are used interchangeably herein. These terms encompass nucleotide sequences and the like. A polynucleotide may be a polymer of RNA or DNA that is single- or double-stranded, that optionally contains synthetic, non-natural or altered nucleotide bases. A polynucleotide in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA, synthetic DNA, or mixtures thereof. An isolated polynucleotide of the present invention may include preferably at least 30 contiguous nucleotides, more preferably at least 40 contiguous nucleotides, most preferably at least 60 contiguous nucleotides derived from SEQ ID NO:3, 5, 7, 11, 13 or 15, or the complement of such sequences.

The term "isolated" refers to materials, such as nucleic acid molecules and/or proteins, which are substantially free or otherwise removed from components that normally accompany or interact with the materials in a naturally occurring environment. Isolated polynucleotides may be purified from a host cell in which they naturally occur. Conventional nucleic acid purification methods known to skilled artisans may be used to obtain isolated polynucleotides. The term also embraces recombinant polynucleotides and chemically synthesized polynucleotides.

The term "recombinant" means, for example, that a nucleic acid sequence is made by an artificial combination of two otherwise separated segments of sequence, e.g., by chemical synthesis or by the manipulation of isolated nucleic acids by genetic engineering techniques. A "recombinant DNA construct" comprises any of the isolated polynucleotides of the present invention operably linked to at least one regulatory sequence.

As used herein, "contig" refers to a nucleotide sequence that is assembled from two or more constituent nucleotide sequences that share common or overlapping regions of sequence homology. For example, the nucleotide sequences of two or more nucleic acid fragments can be compared and aligned in order to identify common or overlapping sequences. Where common or overlapping sequences exist between two or more nucleic acid fragments, the

sequences (and thus their corresponding nucleic acid fragments) can be assembled into a single contiguous nucleotide sequence.

As used herein, "substantially similar" refers to nucleic acid fragments wherein changes in one or more nucleotide bases results in substitution of one or more amino acids, but do not affect the functional properties of the polypeptide encoded by the nucleotide sequence. "Substantially similar" also refers to nucleic acid fragments wherein changes in one or more nucleotide bases does not affect the ability of the nucleic acid fragment to mediate alteration of gene expression by gene silencing through for example antisense or co-suppression technology. "Substantially similar" also refers to modifications of the nucleic acid fragments of the instant invention such as deletion or insertion of one or more nucleotides that do not substantially affect the functional properties of the resulting transcript vis-à-vis the ability to mediate gene silencing or alteration of the functional properties of the resulting protein molecule. It is therefore understood that the invention encompasses more than the specific exemplary nucleotide or amino acid sequences and includes functional equivalents thereof. The terms "substantially similar" and "corresponding substantially" are used interchangeably herein.

The "Clustal V method of alignment" corresponds to the alignment method labeled Clustal V (described by Higgins and Sharp (1989) *CAB/OS*. 5:151-153) and found in the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). The "default parameters" are the parameters pre-set by the manufacturer of the program and for multiple alignments they correspond to GAP PENALTY=10 and GAP LENGTH PENALTY=10, while for pairwise alignments they are KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. After alignment of the sequences, using the Clustal V program, it is possible to obtain a "percent identity" by viewing the "sequence distances" table on the same program.

Substantially similar nucleic acid fragments may be selected by screening nucleic acid fragments representing subfragments or modifications of the nucleic acid fragments of the instant invention, wherein one or more nucleotides are substituted, deleted and/or inserted, for their ability to affect the level of the polypeptide encoded by the unmodified nucleic acid fragment in a plant or plant cell. For example, a substantially similar nucleic acid fragment representing preferably at least 30 contiguous nucleotides, more preferably at least 40 contiguous nucleotides, most preferably at least 60 contiguous nucleotides derived from the instant nucleic acid fragment can be constructed and introduced into a plant or plant cell. The level of the polypeptide encoded by the unmodified nucleic acid fragment present in a

plant or plant cell exposed to the substantially similar nucleic fragment can then be compared to the level of the polypeptide in a plant or plant cell that is not exposed to the substantially similar nucleic acid fragment.

For example, it is well known in the art that antisense suppression and co-suppression of gene expression may be accomplished using nucleic acid fragments representing less than the entire coding region of a gene, and by using nucleic acid fragments that do not share 100% sequence identity with the gene to be suppressed. Moreover, alterations in a nucleic acid fragment which result in the production of a chemically equivalent amino acid at a given site, but do not effect the functional properties of the encoded polypeptide, are well known in the art. Thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also be expected to produce a functionally equivalent product. Nucleotide changes which result in alteration of the N-terminal and C-terminal portions of the polypeptide molecule would also not be expected to alter the activity of the polypeptide. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products. Consequently, an isolated polynucleotide comprising a nucleotide sequence of preferably at least 30 (more preferably at least 40, most preferably at least 60) contiguous nucleotides derived from a nucleotide sequence of SEQ ID NO:3, 5, 7, 11, 13 or 15, and the complement of such nucleotide sequences may be used to affect the expression and/or function of a dihydroflavonol-4-reductase in a host cell. A method of using an isolated polynucleotide to affect the level of expression of a polypeptide in a host cell (eukaryotic, such as plant or yeast, prokaryotic such as bacterial) may comprise the steps of: constructing an isolated polynucleotide of the present invention or an isolated recombinant DNA construct of the present invention; introducing the isolated polynucleotide or the isolated recombinant DNA construct into a host cell; measuring the level of a polypeptide or enzyme activity in the host cell containing the isolated polynucleotide; and comparing the level of a polypeptide or enzyme activity in the host cell containing the isolated polynucleotide with the level of a polypeptide or enzyme activity in a host cell that does not contain the isolated polynucleotide.

Moreover, substantially similar nucleic acid fragments may also be characterized by their ability to hybridize. Estimates of such homology are provided by either DNA-DNA or DNA-RNA hybridization under conditions of stringency as is well understood by those skilled in the art (Hames and Higgins, Eds. (1985) *Nucleic Acid Hybridisation*, IRL Press, Oxford, U.K.). Stringency conditions can be adjusted to screen for moderately similar fragments, such as homologous sequences from distantly related organisms, to highly similar fragments, such as genes that duplicate functional enzymes from closely related organisms. Post-hybridization washes determine stringency conditions. One set of preferred conditions uses a series of washes starting with 6X SSC, 0.5% SDS at room temperature for 15 min, then repeated with 2X SSC, 0.5% SDS at 45 °C for 30 min, and then repeated twice with 0.2X SSC, 0.5% SDS at 50 °C for 30 min. A more preferred set of stringent conditions uses higher temperatures in which the washes are identical to those above except for the temperature of the final two 30 min washes in 0.2X SSC, 0.5% SDS was increased to 60 °C. Another preferred set of highly stringent conditions uses two final washes in 0.1X SSC, 0.1% SDS at 65 °C.

Substantially similar nucleic acid fragments of the instant invention may also be characterized by the percent identity of the amino acid sequences that they encode to the amino acid sequences disclosed herein, as determined by algorithms commonly employed by those skilled in this art. Suitable nucleic acid fragments (isolated polynucleotides of the present invention) encode polypeptides that are at least 70% identical, preferably at least 80% identical to the amino acid sequences reported herein. Preferred nucleic acid fragments encode amino acid sequences that are at least 85% identical to the amino acid sequences reported herein. More preferred nucleic acid fragments encode amino acid sequences that are at least 90% identical to the amino acid sequences reported herein. Most preferred are nucleic acid fragments that encode amino acid sequences that are at least 95% identical to the amino acid sequences reported herein. Suitable nucleic acid fragments not only have the above identities but typically encode a polypeptide having preferably at least 50 amino acids, more preferably at least 100 amino acids, more preferably at least 150 amino acids, still more preferably at least 200 amino acids, and most preferably at least 250 amino acids.

It is well understood by one skilled in the art that many levels of sequence identity are useful in identifying related polypeptide sequences. Useful examples of percent identities are 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%, or any integer percentage from 55% to 100%. Sequence alignments and percent identity calculations were performed using the Megalign program of the

LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal V method of alignment (Higgins and Sharp (1989) *CAB/OS*. 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal V method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

A "substantial portion" of an amino acid or nucleotide sequence comprises an amino acid or a nucleotide sequence that is sufficient to afford putative identification of the protein or gene that the amino acid or nucleotide sequence comprises. Amino acid and nucleotide sequences can be evaluated either manually by one skilled in the art, or by using computer-based sequence comparison and identification tools that employ algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul et al. (1993) *J. Mol. Biol.* 215:403-410; see also the explanation of the BLAST algorithm on the world wide web site for the National Center for Biotechnology Information at the National Library of Medicine of the National Institutes of Health). In general, a sequence of ten or more contiguous amino acids or thirty or more contiguous nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene-specific oligonucleotide probes comprising 30 or more contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., *in situ* hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12 or more nucleotides may be used as amplification primers in PCR in order to obtain a particular nucleic acid fragment comprising the primers. Accordingly, a "substantial portion" of a nucleotide sequence comprises a nucleotide sequence that will afford specific identification and/or isolation of a nucleic acid fragment comprising the sequence. The instant specification teaches amino acid and nucleotide sequences encoding polypeptides that comprise one or more particular plant proteins. The skilled artisan, having the benefit of the sequences as reported herein, may now use all or a substantial portion of the disclosed sequences for purposes known to those skilled in this art. Accordingly, the instant invention comprises the complete sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above.

"Codon degeneracy" refers to divergence in the genetic code permitting variation of the nucleotide sequence without effecting the amino acid sequence of an encoded polypeptide. Accordingly, the instant invention relates to any nucleic

acid fragment comprising a nucleotide sequence that encodes all or a substantial portion of the amino acid sequences set forth herein. The skilled artisan is well aware of the "codon-bias" exhibited by a specific host cell in usage of nucleotide codons to specify a given amino acid. Therefore, when synthesizing a nucleic acid fragment for improved expression in a host cell, it is desirable to design the nucleic acid fragment such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

"Synthetic nucleic acid fragments" can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form larger nucleic acid fragments which may then be enzymatically assembled to construct the entire desired nucleic acid fragment. "Chemically synthesized", as related to a nucleic acid fragment, means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of nucleic acid fragments may be accomplished using well established procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. Accordingly, the nucleic acid fragments can be tailored for optimal gene expression based on optimization of the nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon usage is biased towards those codons favored by the host. Determination of preferred codons can be based on a survey of genes derived from the host cell where sequence information is available.

"Gene" refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. "Native gene" refers to a gene as found in nature with its own regulatory sequences. "Chimeric gene" refers any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. "Endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign-gene" refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, recombinant DNA constructs, or chimeric genes. A "transgene" is an isolated nucleic acid fragment or recombinant

DNA construct that has been introduced into the genome by a transformation procedure.

"Coding sequence" refers to a nucleotide sequence that codes for a specific amino acid sequence. "Regulatory sequences" refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, and polyadenylation recognition sequences.

"Promoter" refers to a nucleotide sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. The promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an "enhancer" is a nucleotide sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. Promoters may be derived in their entirety from a native gene, or may be composed of different elements derived from different promoters found in nature, or may even comprise synthetic nucleotide segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. Promoters which cause a nucleic acid fragment to be expressed in most cell types at most times are commonly referred to as "constitutive promoters". New promoters of various types useful in plant cells are constantly being discovered; numerous examples may be found in the compilation by Okamuro and Goldberg (1989) *Biochemistry of Plants* 15:1-82. It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, nucleic acid fragments of different lengths may have identical promoter activity.

"Translation leader sequence" refers to a nucleotide sequence located between the promoter sequence of a gene and the coding sequence. The translation leader sequence is present in the fully processed mRNA upstream of the translation start sequence. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency. Examples of translation leader sequences have been described (Turner and Foster (1995) *Mol. Biotechnol.* 3:225-236).

"3' non-coding sequences" refer to nucleotide sequences located downstream of a coding sequence and include polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht et al. (1989) *Plant Cell* 1:671-680.

"RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA (mRNA)" refers to the RNA that is without introns and that can be translated into polypeptides by the cell. "cDNA" refers to DNA that is complementary to and derived from an mRNA template. The cDNA can be single-stranded or converted to double stranded form using, for example, the Klenow fragment of DNA polymerase I. "Sense-RNA" refers to an RNA transcript that includes the mRNA and so can be translated into a polypeptide by the cell. "Antisense RNA" refers to an RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene (see U.S. Patent No. 5,107,065, incorporated herein by reference). The complementarity of an antisense RNA may be with any part of the specific nucleotide sequence, i.e., at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. "Functional RNA" refers to sense RNA, antisense RNA, ribozyme RNA, or other RNA that may not be translated but yet has an effect on cellular processes.

The term "operably linked" refers to the association of two or more nucleic acid fragments on a single polynucleotide so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term "expression", as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the invention. Expression may also refer to translation of mRNA into a polypeptide.

"Antisense inhibition" refers to the production of antisense RNA transcripts capable of suppressing the expression of the target protein. "Co-suppression"

refers to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially similar foreign or endogenous genes (U.S. Patent No. 5,231,020, incorporated herein by reference). Co-suppression constructs in plants previously have been designed by focusing on overexpression of a nucleic acid sequence having homology to an endogenous mRNA, in the sense orientation, which results in the reduction of all RNA having homology to the overexpressed sequence (see Vaucheret et al. (1998) *Plant J.* 16:651-659; and Gura (2000) *Nature* 404:804-808). The overall efficiency of this phenomenon is low, and the extent of the RNA reduction is widely variable. Recent work has described the use of "hairpin" structures that incorporate all, or part, of an mRNA encoding sequence in a complementary orientation that results in a potential "stem-loop" structure for the expressed RNA (PCT Publication WO 99/53050 published on October 21, 1999 and more recently, Applicants' assignee's own WO 02/00904 published on January 3, 2002). This increases the frequency of co-suppression in the recovered transgenic plants. Hairpin structures may contain the target RNA forming either the stem or the loop of the hairpin. Another variation describes the use of plant viral sequences to direct the suppression, or "silencing", of proximal mRNA encoding sequences (PCT Publication WO 98/36083 published on August 20, 1998). Both of these co-suppressing phenomena have not been elucidated mechanistically, although genetic evidence has begun to unravel this complex situation (Elmayan et al. (1998) *Plant Cell* 10:1747-1757).

"Overexpression" refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or non-transformed organisms.

A "protein" or "polypeptide" is a chain of amino acids arranged in a specific order determined by the coding sequence in a polynucleotide encoding the polypeptide. Each protein or polypeptide has a unique function.

"Altered levels" or "altered expression" refers to the production of gene product(s) in transgenic organisms in amounts or proportions that differ from that of normal or non-transformed organisms.

"Mature protein" or the term "mature" when used in describing a protein refers to a post-translationally processed polypeptide; i.e., one from which any pre- or propeptides present in the primary translation product have been removed. "Precursor protein" or the term "precursor" when used in describing a protein refers to the primary product of translation of mRNA; i.e., with pre- and propeptides still present. Pre- and propeptides may be but are not limited to intracellular localization signals.

A "chloroplast transit peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the chloroplast or other plastid types present in the cell in which the protein is made. "Chloroplast transit sequence" refers to a nucleotide sequence that encodes a chloroplast transit peptide. A "signal peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the secretory system (Chrispeels (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol.* 42:21-53). If the protein is to be directed to a vacuole, a vacuolar targeting signal (*supra*) can further be added, or if to the endoplasmic reticulum, an endoplasmic reticulum retention signal (*supra*) may be added. If the protein is to be directed to the nucleus, any signal peptide present should be removed and instead a nuclear localization signal included (Raikhel (1992) *Plant Phys.* 100:1627-1632).

"Transformation" refers to the transfer of a nucleic acid fragment into the genome of a host organism. Host organisms containing the transferred nucleic acid fragments are referred to as "transgenic" or "transformed" organisms. Examples of methods of plant transformation include *Agrobacterium*-mediated transformation (De Blaere et al. (1987) *Meth. Enzymol.* 143:277; Ishida Y. et al. (1996) *Nature Biotech.* 14:745-750) and particle-accelerated or "gene gun" transformation technology (Klein et al. (1987) *Nature (London)* 327:70-73; U.S. Patent No. 4,945,050, incorporated herein by reference). Thus, isolated polynucleotides of the present invention can be incorporated into recombinant constructs, typically DNA constructs, capable of introduction into and replication in a host cell. Such a construct can be a vector that includes a replication system and sequences that are capable of transcription and translation of a polypeptide-encoding sequence in a given host cell. A number of vectors suitable for stable transfection of plant cells or for the establishment of transgenic plants have been described in, e.g., Pouwels et al., *Cloning Vectors: A Laboratory Manual*, 1985, supp. 1987; Weissbach and Weissbach, *Methods for Plant Molecular Biology*, Academic Press, 1989; and Flevin et al., *Plant Molecular Biology Manual*, Kluwer Academic Publishers, 1990. Typically, plant expression vectors include, for example, one or more cloned plant genes under the transcriptional control of 5' and 3' regulatory sequences and a dominant selectable marker. Such plant expression vectors also can contain a promoter regulatory region (e.g., a regulatory region controlling inducible or constitutive, environmentally- or developmentally-regulated, or cell- or tissue-specific expression), a transcription initiation start site, a ribosome binding site, an RNA processing signal, a transcription termination site, and/or a polyadenylation signal.

"Stable transformation" refers to the transfer of a nucleic acid fragment into a genome of a host organism, including both nuclear and organellar genomes, resulting in genetically stable inheritance. In contrast, "transient transformation" refers to the transfer of a nucleic acid fragment into the nucleus, or DNA-containing organelle, of a host organism resulting in gene expression without integration or stable inheritance. The term "transformation" as used herein refers to both stable transformation and transient transformation.

The terms "recombinant construct", "expression construct" and "recombinant expression construct" are used interchangeably herein. These terms refer to a functional unit of genetic material that can be inserted into the genome of a cell using standard methodology well known to one skilled in the art. Such construct may be used by itself or may be used in conjunction with a vector. If a vector is used, the choice of vector is dependent upon the method that will be used to transform host plants as is well known to those skilled in the art. .

Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described more fully in Sambrook et al. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1989 (hereinafter "Maniatis").

"Motifs" or "subsequences" refer to short regions of conserved sequences of nucleic acids or amino acids that comprise part of a longer sequence. For example, it is expected that such conserved subsequences would be important for function, and could be used to identify new homologues in plants. It is expected that some or all of the elements may be found in a homologue. Also, it is expected that one or two of the conserved amino acids in any given motif may differ in a true homologue.

"PCR" or "polymerase chain reaction" is well known by those skilled in the art as a technique used for the amplification of specific DNA segments (U.S. Patent Nos. 4,683,195 and 4,800,159).

The present invention concerns an isolated polynucleotide comprising a nucleotide sequence encoding a dihydroflavonol-4-reductase polypeptide having at least 80% identity, based on the Clustal V method of alignment, when compared to one of SEQ ID NO:4, 6, 8, 12, 14, or 16.

This invention also includes to the isolated complement of such polynucleotides, wherein the complement and the polynucleotide consist of the same number of nucleotides, and the nucleotide sequences of the complement and the polynucleotide have 100% complementarity.

Nucleic acid fragments encoding at least a portion of several dihydroflavonol-4-reductase proteins have been isolated and identified by comparison of random

plant cDNA sequences to public databases containing nucleotide and protein sequences using the BLAST algorithms well known to those skilled in the art. The nucleic acid fragments of the instant invention may be used to isolate cDNAs and genes encoding homologous proteins from the same or other plant species. Isolation of homologous genes using sequence-dependent protocols is well known in the art. Examples of sequence-dependent protocols include, but are not limited to, methods of nucleic acid hybridization, and methods of DNA and RNA amplification as exemplified by various uses of nucleic acid amplification technologies (e.g., polymerase chain reaction, ligase chain reaction).

For example, genes encoding other dihydroflavonol-4-reductase, either as cDNAs or genomic DNAs, could be isolated directly by using all or a portion of the instant nucleic acid fragments as DNA hybridization probes to screen libraries from any desired plant employing methodology well known to those skilled in the art. Specific oligonucleotide probes based upon the instant nucleic acid sequences can be designed and synthesized by methods known in the art (Maniatis). Moreover, an entire sequence can be used directly to synthesize DNA probes by methods known to the skilled artisan such as random primer DNA labeling, nick translation, end-labeling techniques, or RNA probes using available *in vitro* transcription systems. In addition, specific primers can be designed and used to amplify a part or all of the instant sequences. The resulting amplification products can be labeled directly during amplification reactions or labeled after amplification reactions, and used as probes to isolate full length cDNA or genomic fragments under conditions of appropriate stringency.

In addition, two short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols to amplify longer nucleic acid fragments encoding homologous genes from DNA or RNA. The polymerase chain reaction may also be performed on a library of cloned nucleic acid fragments wherein the sequence of one primer is derived from the instant nucleic acid fragments, and the sequence of the other primer takes advantage of the presence of the polyadenylic acid tracts to the 3' end of the mRNA precursor encoding plant genes. Alternatively, the second primer sequence may be based upon sequences derived from the cloning vector. For example, the skilled artisan can follow the RACE protocol (Frohman et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:8998-9002) to generate cDNAs by using PCR to amplify copies of the region between a single point in the transcript and the 3' or 5' end. Primers oriented in the 3' and 5' directions can be designed from the instant sequences. Using commercially available 3' RACE or 5' RACE systems (BRL), specific 3' or 5' cDNA fragments can

be isolated (Ohara et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:5673-5677; Loh et al. (1989) *Science* 243:217-220). Products generated by the 3' and 5' RACE procedures can be combined to generate full-length cDNAs (Frohman and Martin (1989) *Techniques* 1:165). Consequently, a polynucleotide comprising a nucleotide sequence of preferably at least 30 (more preferably at least 40, most preferably at least 60) contiguous nucleotides derived from a nucleotide sequence of SEQ ID NOs:3, 5, 7, 11, 13 or 15 and the complement of such nucleotide sequences may be used in such methods to obtain a nucleic acid fragment encoding a substantial portion of an amino acid sequence of a polypeptide.

Availability of the instant nucleotide and deduced amino acid sequences facilitates immunological screening of cDNA expression libraries. Synthetic peptides representing portions of the instant amino acid sequences may be synthesized. These peptides can be used to immunize animals to produce polyclonal or monoclonal antibodies with specificity for peptides or proteins comprising the amino acid sequences. These antibodies can be then be used to screen cDNA expression libraries to isolate full-length cDNA clones of interest (Lerner (1984) *Adv. Immunol.* 36:1-34; Maniatis).

In another embodiment, this invention concerns viruses and host cells comprising either the recombinant DNA constructs of the invention as described herein or isolated polynucleotides of the invention as described herein. Examples of host cells which can be used to practice the invention include, but are not limited to, yeast, bacteria, and plants.

As was noted above, the nucleic acid fragments of the instant invention may be used to create transgenic plants in which the disclosed polypeptides are present at higher or lower levels than normal or in cell types or developmental stages in which they are not normally found. This would have the effect of altering the level of compounds found within the phenylpropanoid pathway in those cells. Furthermore, it is believed that the polynucleotides of the instant invention may be used to create transgenic plants where the dihydroflavonol-4-reductase levels are altered with respect to non-transgenic plants which would result in plants with an altered phenotype. The use of fragments encoding dihydroflavonol-4-reductase for altering the levels of flavonoids has been previously suggested. U.S. Patent No. 5,410,096 and U.S. Patent No. 6,080,920 discuss methods of generating transgenic plants with altered/modified flower color. Since flavonoids are involved in flower and fruit color, manipulation of the dihydroflavonol-4-reductase levels may also affect the color of the seed produced.

Overexpression of the polypeptides of the instant invention may be accomplished by first constructing a recombinant DNA construct in which the coding region is operably linked to a promoter capable of directing expression of a polypeptide in the desired tissues at the desired stage of development. The recombinant DNA construct may comprise promoter sequences and translation leader sequences derived from the same genes. 3' Non-coding sequences corresponding to transcription termination signals may also be provided. The instant recombinant DNA construct may also comprise one or more introns in order to facilitate expression of the recombinant DNA construct.

Plasmid vectors comprising the instant isolated polynucleotide(s) (or recombinant DNA construct(s)) may be constructed. The choice of plasmid vector is dependent upon the method that will be used to transform host plants. The skilled artisan is well aware of the genetic elements that must be present on the plasmid vector in order to successfully transform, select and propagate host cells containing the recombinant DNA construct or chimeric gene. The skilled artisan will also recognize that different independent transformation events will result in different levels and patterns of expression (Jones et al. (1985) *EMBO J.* 4:2411-2418; De Almeida et al. (1989) *Mol. Gen. Genetics* 218:78-86), and thus that multiple events must be screened in order to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished by Southern analysis of DNA, Northern analysis of mRNA expression, Western analysis of protein expression, or phenotypic analysis.

For some applications it may be useful to direct the instant polypeptides to different cellular compartments, or to facilitate its secretion from the cell. It is thus envisioned that the recombinant DNA construct(s) described above may be further supplemented by directing the coding sequence to encode the instant polypeptides with appropriate intracellular targeting sequences such as transit sequences (Keegstra (1989) *Cell* 56:247-253), signal sequences or sequences encoding endoplasmic reticulum localization (Chrispeels (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol.* 42:21-53), or nuclear localization signals (Raikhel (1992) *Plant Phys.* 100:1627-1632) with or without removing targeting sequences that are already present. While the references cited give examples of each of these, the list is not exhaustive and more targeting signals of use may be discovered in the future.

It may also be desirable to reduce or eliminate expression of genes encoding the instant polypeptides in plants for some applications. In order to accomplish this, a recombinant DNA construct designed for co-suppression of the instant polypeptide can be constructed by linking a gene or gene fragment encoding that

polypeptide to plant promoter sequences. Alternatively, a recombinant DNA construct designed to express antisense RNA for all or part of the instant nucleic acid fragment can be constructed by linking the gene or gene fragment in reverse orientation to plant promoter sequences. Either the co-suppression or antisense recombinant DNA constructs could be introduced into plants via transformation wherein expression of the corresponding endogenous genes are reduced or eliminated.

Molecular genetic solutions to the generation of plants with altered gene expression have a decided advantage over more traditional plant breeding approaches. Changes in plant phenotypes can be produced by specifically inhibiting expression of one or more genes by antisense inhibition or cosuppression (U.S. Patent Nos. 5,190,931, 5,107,065 and 5,283,323). An antisense or cosuppression construct would act as a dominant negative regulator of gene activity. While conventional mutations can yield negative regulation of gene activity these effects are most likely recessive. The dominant negative regulation available with a transgenic approach may be advantageous from a breeding perspective. In addition, the ability to restrict the expression of a specific phenotype to the reproductive tissues of the plant by the use of tissue specific promoters may confer agronomic advantages relative to conventional mutations which may have an effect in all tissues in which a mutant gene is ordinarily expressed.

The person skilled in the art will know that special considerations are associated with the use of antisense or cosuppression technologies in order to reduce expression of particular genes. For example, the proper level of expression of sense or antisense genes may require the use of different recombinant DNA constructs utilizing different regulatory elements known to the skilled artisan. Once transgenic plants are obtained by one of the methods described above, it will be necessary to screen individual transgenics for those that most effectively display the desired phenotype. Accordingly, the skilled artisan will develop methods for screening large numbers of transformants. The nature of these screens will generally be chosen on practical grounds. For example, one can screen by looking for changes in gene expression by using antibodies specific for the protein encoded by the gene being suppressed, or one could establish assays that specifically measure enzyme activity. A preferred method will be one which allows large numbers of samples to be processed rapidly, since it will be expected that a large number of transformants will be negative for the desired phenotype.

In a preferred embodiment, a dihydroflavonol-4-reductase polypeptide has an amino acid sequence that is at least 80% identical, based on the Clustal V method of alignment, to a polypeptide of SEQ ID NO:4, 6, 8, 12, 14, or 16.

The instant polypeptides (or portions thereof) may be produced in heterologous host cells, particularly in the cells of microbial hosts, and can be used to prepare antibodies to these proteins by methods well known to those skilled in the art. The antibodies are useful for detecting the polypeptides of the instant invention *in situ* in cells or *in vitro* in cell extracts. Preferred heterologous host cells for production of the instant polypeptides are microbial hosts. Microbial expression systems and expression vectors containing regulatory sequences that direct high level expression of foreign proteins are well known to those skilled in the art. Any of these could be used to construct a recombinant DNA construct for production of the instant polypeptides. This recombinant DNA construct could then be introduced into appropriate microorganisms via transformation to provide high level expression of the encoded dihydroflavonol-4-reductase. An example of a vector for high level expression of the instant polypeptides in a bacterial host is provided (Example 6).

All or a substantial portion of the polynucleotides of the instant invention may also be used as probes for genetically and physically mapping the genes that they are a part of, and used as markers for traits linked to those genes. Such information may be useful in plant breeding in order to develop lines with desired phenotypes. For example, the instant nucleic acid fragments may be used as restriction fragment length polymorphism (RFLP) markers. Southern blots (Maniatis) of restriction-digested plant genomic DNA may be probed with the nucleic acid fragments of the instant invention. The resulting banding patterns may then be subjected to genetic analyses using computer programs such as MapMaker (Lander et al. (1987) *Genomics* 1:174-181) in order to construct a genetic map. In addition, the nucleic acid fragments of the instant invention may be used to probe Southern blots containing restriction endonuclease-treated genomic DNAs of a set of individuals representing parent and progeny of a defined genetic cross. Segregation of the DNA polymorphisms is noted and used to calculate the position of the instant nucleic acid sequence in the genetic map previously obtained using this population (Botstein et al. (1980) *Am. J. Hum. Genet.* 32:314-331).

The production and use of plant gene-derived probes for use in genetic mapping is described in Bernatzky and Tanksley (1986) *Plant Mol. Biol. Reporter* 4:37-41. Numerous publications describe genetic mapping of specific cDNA clones using the methodology outlined above or variations thereof. For example, F₂ intercross populations, backcross populations, randomly mated populations, near

isogenic lines, and other sets of individuals may be used for mapping. Such methodologies are well known to those skilled in the art.

Nucleic acid probes derived from the instant nucleic acid sequences may also be used for physical mapping (i.e., placement of sequences on physical maps; see Hoheisel et al. In: *Nonmammalian Genomic Analysis: A Practical Guide*, Academic press 1996, pp. 319-346, and references cited therein).

Nucleic acid probes derived from the instant nucleic acid sequences may be used in direct fluorescence *in situ* hybridization (FISH) mapping (Trask (1991) *Trends Genet.* 7:149-154). Although current methods of FISH mapping favor use of large clones (several kb to several hundred kb; see Laan et al. (1995) *Genome Res.* 5:13-20), improvements in sensitivity may allow performance of FISH mapping using shorter probes.

A variety of nucleic acid amplification-based methods of genetic and physical mapping may be carried out using the instant nucleic acid sequences. Examples include allele-specific amplification (Kazazian (1989) *J. Lab. Clin. Med.* 11:95-96), polymorphism of PCR-amplified fragments (CAPS; Sheffield et al. (1993) *Genomics* 16:325-332), allele-specific ligation (Landegren et al. (1988) *Science* 241:1077-1080), nucleotide extension reactions (Sokolov (1990) *Nucleic Acid Res.* 18:3671), Radiation Hybrid Mapping (Walter et al. (1997) *Nat. Genet.* 7:22-28) and Happy Mapping (Dear and Cook (1989) *Nucleic Acid Res.* 17:6795-6807). For these methods, the sequence of a nucleic acid fragment is used to design and produce primer pairs for use in the amplification reaction or in primer extension reactions. The design of such primers is well known to those skilled in the art. In methods employing PCR-based genetic mapping, it may be necessary to identify DNA sequence differences between the parents of the mapping cross in the region corresponding to the instant nucleic acid sequence. This, however, is generally not necessary for mapping methods.

Loss of function mutant phenotypes may be identified for the instant cDNA clones either by targeted gene disruption protocols or by identifying specific mutants for these genes contained in a maize population carrying mutations in all possible genes (Ballinger and Benzer (1989) *Proc. Natl. Acad. Sci USA* 86:9402-9406; Koes et al. (1995) *Proc. Natl. Acad. Sci USA* 92:8149-8153; Bensen et al. (1995) *Plant Cell* 7:75-84). The latter approach may be accomplished in two ways. First, short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols in conjunction with a mutation tag sequence primer on DNAs prepared from a population of plants in which Mutator transposons or some other mutation-causing DNA element has been introduced (see Bensen, *supra*). The

amplification of a specific DNA fragment with these primers indicates the insertion of the mutation tag element in or near the plant gene encoding the instant polypeptides. Alternatively, the instant nucleic acid fragment may be used as a hybridization probe against PCR amplification products generated from the mutation population using the mutation tag sequence primer in conjunction with an arbitrary genomic site primer, such as that for a restriction enzyme site-anchored synthetic adaptor. With either method, a plant containing a mutation in the endogenous gene encoding the instant polypeptides can be identified and obtained. This mutant plant can then be used to determine or confirm the natural function of the instant polypeptides disclosed herein.

EXAMPLES

The present invention is further defined in the following Examples, in which parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions. Thus, various modifications of the invention in addition to those shown and described herein will be apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the appended claims.

EXAMPLE 1

Composition of cDNA Libraries:

Isolation and Sequencing of cDNA Clones

cDNA libraries representing mRNAs from various soybean (*Glycine max*) and wheat (*Triticum aestivum*) tissues were prepared. The characteristics of the libraries are described below.

TABLE 2

cDNA Libraries from Soybean and Wheat

Library	Tissue	Clone
ssl1c	Soybean (<i>Glycine max</i>), Transgenic High Lysine Line 5403-218 Seed 25 Days After Flowering	ssl1c.pk004.m3
sre	Soybean (<i>Glycine max</i>) Root Elongation Zone 4 to 5 Days After Germination	sre.pk0043.d11:fis
sls2c	Soybean (<i>Glycine max</i>) Infected With <i>Sclerotinia</i>	sls2c.pk027.c22:fis

	<i>sclerotiorum</i> Mycelium	
ssl1c	Soybean (<i>Glycine max</i>) Transgenic High Lysine Line ssl1c.pk004.m3:fis 5403-218 Seed 25 Days After Flowering	
wdi1c	Wheat (<i>Triticum aestivum</i> , Hi Line) developing inflorescence +/- 4 cm	wdi1c.pk006.d20
wdk3c	Wheat (<i>Triticum aestivum</i>) Developing Kernel, 14 Days After Anthesis	wdk3c.pk012.j12:fis
wdi1c	Wheat (<i>Triticum aestivum</i> , Hi Line) developing inflorescence +/- 4 cm	wdi1c.pk006.d20:fis
wdk2c	Wheat (<i>Triticum aestivum</i>) Developing Kernel, 7 Days After Anthesis	wdk2c.pk017.h12:fis

cDNA libraries may be prepared by any one of many methods available. For example, the cDNAs may be introduced into plasmid vectors by first preparing the cDNA libraries in Uni-ZAP™ XR vectors according to the manufacturer's protocol (Stratagene Cloning Systems, La Jolla, CA). The Uni-ZAP™ XR libraries are converted into plasmid libraries according to the protocol provided by Stratagene. Upon conversion, cDNA inserts will be contained in the plasmid vector pBluescript. In addition, the cDNAs may be introduced directly into precut Bluescript II SK(+) vectors (Stratagene) using T4 DNA ligase (New England Biolabs), followed by transfection into DH10B cells according to the manufacturer's protocol (GIBCO BRL Products). Once the cDNA inserts are in plasmid vectors, plasmid DNAs are prepared from randomly picked bacterial colonies containing recombinant pBluescript plasmids, or the insert cDNA sequences are amplified via polymerase chain reaction using primers specific for vector sequences flanking the inserted cDNA sequences. Amplified insert DNAs or plasmid DNAs are sequenced in dye-primer sequencing reactions to generate partial cDNA sequences (expressed sequence tags or "ESTs"; see Adams et al., (1991) *Science* 252:1651-1656). The resulting ESTs are analyzed using a Perkin Elmer Model 377 fluorescent sequencer.

Full-insert sequence (FIS) data is generated utilizing a modified transposition protocol. Clones identified for FIS are recovered from archived glycerol stocks as single colonies, and plasmid DNAs are isolated via alkaline lysis. Isolated DNA templates are reacted with vector primed M13 forward and reverse oligonucleotides in a PCR-based sequencing reaction and loaded onto automated sequencers. Confirmation of clone identification is performed by sequence alignment to the original EST sequence from which the FIS request is made.

Confirmed templates are transposed via the Primer Island transposition kit (PE Applied Biosystems, Foster City, CA) which is based upon the *Saccharomyces cerevisiae* Ty1 transposable element (Devine and Boeke (1994) *Nucleic Acids Res.* 22:3765-3772). The *in vitro* transposition system places unique binding sites randomly throughout a population of large DNA molecules. The transposed DNA is then used to transform DH10B electro-competent cells (Gibco BRL/Life Technologies, Rockville, MD) via electroporation. The transposable element contains an additional selectable marker (named DHFR; Fling and Richards (1983) *Nucleic Acids Res.* 11:5147-5158), allowing for dual selection on agar plates of only those subclones containing the integrated transposon. Multiple subclones are randomly selected from each transposition reaction, plasmid DNAs are prepared via alkaline lysis, and templates are sequenced (ABI Prism dye-terminator ReadyReaction mix) outward from the transposition event site, utilizing unique primers specific to the binding sites within the transposon.

Sequence data is collected (ABI Prism Collections) and assembled using Phred/Phrap (P. Green, University of Washington, Seattle). Phred/Phrap is a public domain software program which re-reads the ABI sequence data, re-calls the bases, assigns quality values, and writes the base calls and quality values into editable output files. The Phrap sequence assembly program uses these quality values to increase the accuracy of the assembled sequence contigs. Assemblies are viewed by the Consed sequence editor (D. Gordon, University of Washington, Seattle).

In some of the clones the cDNA fragment corresponds to a portion of the 3'-terminus of the gene and does not cover the entire open reading frame. In order to obtain the upstream information one of two different protocols are used. The first of these methods results in the production of a fragment of DNA containing a portion of the desired gene sequence while the second method results in the production of a fragment containing the entire open reading frame. Both of these methods use two rounds of PCR amplification to obtain fragments from one or more libraries. The libraries some times are chosen based on previous knowledge that the specific gene should be found in a certain tissue and some times are randomly-chosen. Reactions to obtain the same gene may be performed on several libraries in parallel or on a pool of libraries. Library pools are normally prepared using from 3 to 5 different libraries and normalized to a uniform dilution. In the first round of amplification both methods use a vector-specific (forward) primer corresponding to a portion of the vector located at the 5'-terminus of the clone coupled with a gene-specific (reverse) primer. The first method uses a sequence that is complementary to a portion of the already known gene sequence while the second

method uses a gene-specific primer complementary to a portion of the 3'-untranslated region (also referred to as UTR). In the second round of amplification a nested set of primers is used for both methods. The resulting DNA fragment is ligated into a pBluescript vector using a commercial kit and following the manufacturer's protocol. This kit is selected from many available from several vendors including Invitrogen (Carlsbad, CA), Promega Biotech (Madison, WI), and Gibco-BRL (Gaithersburg, MD). The plasmid DNA is isolated by alkaline lysis method and submitted for sequencing and assembly using Phred/Phrap, as above.

EXAMPLE 2

Identification of cDNA Clones

cDNA clones encoding dihydroflavonol-4-reductase were identified by conducting BLAST (Basic Local Alignment Search Tool; Altschul et al. (1993) *J. Mol. Biol.* 215:403-410; see also the explanation of the BLAST algorithm on the world wide web site for the National Center for Biotechnology Information at the National Library of Medicine of the National Institutes of Health) searches for similarity to sequences contained in the BLAST "nr" database (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the last major release of the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The cDNA sequences obtained in Example 1 were analyzed for similarity to all publicly available DNA sequences contained in the "nr" database using the BLASTN algorithm provided by the National Center for Biotechnology Information (NCBI). The DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database using the BLASTX algorithm (Gish and States (1993) *Nat. Genet.* 3:266-272) provided by the NCBI. For convenience, the P-value (probability) of observing a match of a cDNA sequence to a sequence contained in the searched databases merely by chance as calculated by BLAST are reported herein as "pLog" values, which represent the negative of the logarithm of the reported P-value. Accordingly, the greater the pLog value, the greater the likelihood that the cDNA sequence and the BLAST "hit" represent homologous proteins.

ESTs submitted for analysis are compared to the genbank database as described above. ESTs that contain sequences more 5- or 3-prime can be found by using the BLASTn algorithm (Altschul et al (1997) *Nucleic Acids Res.* 25:3389-3402.) against the Du Pont proprietary database comparing nucleotide sequences that share common or overlapping regions of sequence homology. Where common or overlapping sequences exist between two or more nucleic acid

fragments, the sequences can be assembled into a single contiguous nucleotide sequence, thus extending the original fragment in either the 5 or 3 prime direction. Once the most 5-prime EST is identified, its complete sequence can be determined by Full Insert Sequencing as described in Example 1. Homologous genes belonging to different species can be found by comparing the amino acid sequence of a known gene (from either a proprietary source or a public database) against an EST database using the tBLASTn algorithm. The tBLASTn algorithm searches an amino acid query against a nucleotide database that is translated in all 6 reading frames. This search allows for differences in nucleotide codon usage between different species, and for codon degeneracy.

EXAMPLE 3

Characterization of cDNA Clones Encoding Dihydroflavonol-4-Reductase

The BLASTX search using the EST sequences from clones listed in Table 3 revealed similarity of the polypeptides encoded by the cDNAs to dihydroflavonol-4-reductase from *Glycine max* (NCBI General Identifier No. 5852933; SEQ ID NO:17), *Arabidopsis thaliana* (NCBI General Identifier No. 6689826; SEQ ID NO:18), *Hordeum vulgare* (NCBI General Identifier No. 1706372; SEQ ID NO:19), *Oryza sativa* (NCBI General Identifier No. 4581047; SEQ ID NO:20), *Fragaria ananassa* (NCBI General Identifier No. 2599562; SEQ ID NO:21), and *Zea mays* (NCBI General Identifier No. 14030554; SEQ ID NO:22). Shown in Table 3 are the BLAST results for individual ESTs ("EST"), the sequences of the entire cDNA inserts comprising the indicated cDNA clones ("FIS"), or sequences encoding an entire protein derived from an FIS, a contig, or an FIS and PCR ("CGS"):

TABLE 3
BLAST Results for Sequences Encoding Polypeptides Homologous
To Dihydroflavonol-4-Reductase

Clone	Status	NCBI General Identifier No. (SEQ ID NO:)	BLAST pLog Score
ssl1c.pk004.m3	EST	2599562 (SEQ ID NO:21)	34.05
sre.pk0043.d11:fis	CGS	5852933 (SEQ ID NO:17)	165.00
sls2c.pk027.c22:fis	CGS	5852933 (SEQ ID NO:17)	164.00

ssl1c.pk004.m3:fis	FIS	6689826 (SEQ ID NO:18)	88.00
wdi1c.pk006.d20	EST	14030554 (SEQ ID NO:22)	79.40
wdk3c.pk012.j12:fis	CGS	1706372 (SEQ ID NO:19)	180.00
wdi1c.pk006.d20:fis	CGS	4581047 (SEQ ID NO:20)	135.00
wdk2c.pk017.h12:fis	CGS	1706372 (SEQ ID NO:19)	180.00

The nucleotide sequence of clone ssl1c.pk004.m3 is shown in SEQ ID NO:1. The amino acid sequence deduced from nucleotides 58 through 420 of SEQ ID NO:1 is shown in SEQ ID NO:2. The nucleotide sequence of clone sre.pk0043.d11:fis is shown in SEQ ID NO:3. The amino acid sequence deduced from nucleotides 43 through 1062 of SEQ ID NO:3 is shown in SEQ ID NO:4 (an ATG start codon at nucleotides 43-45 and a TAA stop codon at nucleotides 1060-1062). The nucleotide sequence of clone sls2c.pk027.c22:fis is shown in SEQ ID NO:5. The amino acid sequence deduced from nucleotides 112 through 1131 of SEQ ID NO:5 is shown in SEQ ID NO:6 (an ATG start codon at nucleotides 112-114 and a TAA stop codon at nucleotides 1129-1131). The nucleotide sequence of clone ssl1c.pk004.m3:fis is shown in SEQ ID NO:7. The amino acid sequence deduced from nucleotides 58 through 771 of SEQ ID NO:7 is shown in SEQ ID NO:8. The nucleotide sequence of clone wdi1c.pk006.d20 is shown in SEQ ID NO:9. The amino acid sequence deduced from nucleotides 15 through 546 of SEQ ID NO:9 is shown in SEQ ID NO:10. The nucleotide sequence of clone wdk3c.pk012.j12:fis is shown in SEQ ID NO:11. The amino acid sequence deduced from nucleotides 30 through 1109 of SEQ ID NO:11 is shown in SEQ ID NO:12 (an ATG start codon at nucleotides 30-32 and a TAA stop codon at nucleotides 1107-1109). The nucleotide sequence of clone wdi1c.pk006.d20:fis is shown in SEQ ID NO:13. The amino acid sequence deduced from nucleotides 15 through 10004 of SEQ ID NO:13 is shown in SEQ ID NO:14 (an ATG start codon at nucleotides 15-17 and a TGA stop codon at nucleotides 1002-1004). The nucleotide sequence of clone wdk2c.pk017.h12:fis is shown in SEQ ID NO:15. The amino acid sequence deduced from nucleotides 48 through 1127 of SEQ ID NO:15 is shown in SEQ ID NO:16 (an ATG start codon at nucleotides 48-50 and a TAA stop codon at nucleotides 1125-1127).

Figures 2A, 2B and 2C present an alignment of the amino acid sequences set forth in SEQ ID NOs:4, 6, 8 and the amino acid sequence of *Glycine max* dihydroflavonol-4-reductase (NCBI General Identifier No. 5852933; SEQ ID NO:17) and *Arabidopsis thaliana* dihydroflavonol-4-reductase (NCBI General Identifier No. 6689826; SEQ ID NO:18). The data in Table 4 includes a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:4, 6, 8 and the amino acid sequence of *Glycine max* dihydroflavonol-4-reductase (NCBI General Identifier No. 5852933; SEQ ID NO:17) or *Arabidopsis thaliana* dihydroflavonol-4-reductase (NCBI General Identifier No. 6689826; SEQ ID NO:18).

Figures 3A and 3B present an alignment of the amino acid sequences set forth in SEQ ID NOs:12, 14, 16 and the amino acid sequence of *Hordeum vulgare* dihydroflavonol-4-reductase (NCBI General Identifier No. 1706372; SEQ ID NO:19) and *Oryza sativa* dihydroflavonol-4-reductase (NCBI General Identifier No. 4581047; SEQ ID NO:20). The data in Table 4 includes a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:12, 14, 16 and the amino acid sequence of *Hordeum vulgare* dihydroflavonol-4-reductase (NCBI General Identifier No. 1706372; SEQ ID NO:19) or *Oryza sativa* dihydroflavonol-4-reductase (NCBI General Identifier No. 4581047; SEQ ID NO:20).

TABLE 4
Percent Identity of Amino Acid Sequences Deduced From the
Nucleotide Sequences of cDNA Clones Encoding Polypeptides
Homologous to Dihydroflavonol-4-Reductase

Clone	SEQ ID NO:	NCBI General Identifier No. (SEQ ID NO:)	Percent Identity
ssl1c.pk004.m3	2	2599562 (SEQ ID NO:21)	46.0
sre.pk0043.d11:fis	4	5852933 (SEQ ID NO:17)	80.2
sls2c.pk027.c22:fis	6	5852933 (SEQ ID NO:17)	79.9
ssl1c.pk004.m3:fis	8	6689826 (SEQ ID NO:18)	60.5
wdi1c.pk006.d20	10	14030554 (SEQ ID NO:22)	70.1
wdk3c.pk012.j12:fis	12	1706372 (SEQ ID NO:19)	95.5

wdi1c.pk006.d20:fis	14	4581047 (SEQ ID NO:20)	70.2
wdk2c.pk017.h12:fis	16	1706372 (SEQ ID NO:19)	94.4

Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal V method of alignment (Higgins and Sharp (1989) *CABIOS*. 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. Sequence alignments and BLAST scores and probabilities indicate that the nucleic acid fragments comprising the instant cDNA clones encode a substantial portion of a dihydroflavonol-4-reductase. These sequences represent the first wheat sequences encoding dihydroflavonol-4-reductase known to Applicant.

EXAMPLE 4

Expression of Recombinant DNA Constructs in Monocot Cells

A recombinant DNA construct comprising a cDNA encoding the instant polypeptides in sense orientation with respect to the maize 27 kD zein promoter that is located 5' to the cDNA fragment, and the 10 kD zein 3' end that is located 3' to the cDNA fragment, can be constructed. The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites (NcoI or SmaI) can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the digested vector pML103 as described below. Amplification is then performed in a standard PCR. The amplified DNA is then digested with restriction enzymes NcoI and SmaI and fractionated on an agarose gel. The appropriate band can be isolated from the gel and combined with a 4.9 kb NcoI-SmaI fragment of the plasmid pML103. Plasmid pML103 has been deposited under the terms of the Budapest Treaty at ATCC (American Type Culture Collection, 10801 University Blvd., Manassas, VA 20110-2209), and bears accession number ATCC 97366. The DNA segment from pML103 contains a 1.05 kb Sall-NcoI promoter fragment of the maize 27 kD zein gene and a 0.96 kb SmaI-Sall fragment from the 3' end of the maize 10 kD zein gene in the vector pGem9Zf(+) (Promega). Vector and insert DNA can be ligated at 15 °C overnight, essentially as described (Maniatis). The ligated DNA may then be used to transform *E. coli* XL1-Blue (Epicurian Coli XL-1 Blue™; Stratagene). Bacterial transformants can be screened by restriction enzyme

digestion of plasmid DNA and limited nucleotide sequence analysis using the dideoxy chain termination method (Sequenase™ DNA Sequencing Kit; U.S. Biochemical). The resulting plasmid construct would comprise a recombinant DNA construct encoding, in the 5' to 3' direction, the maize 27 kD zein promoter, a cDNA fragment encoding the instant polypeptides, and the 10 kD zein 3' region.

The recombinant DNA construct described above can then be introduced into corn cells by the following procedure. Immature corn embryos can be dissected from developing caryopses derived from crosses of the inbred corn lines H99 and LH132. The embryos are isolated 10 to 11 days after pollination when they are 1.0 to 1.5 mm long. The embryos are then placed with the axis-side facing down and in contact with agarose-solidified N6 medium (Chu et al. (1975) *Sci. Sin. Peking* 18:659-668). The embryos are kept in the dark at 27 °C. Friable embryogenic callus consisting of undifferentiated masses of cells with somatic proembryoids and embryoids borne on suspensor structures proliferates from the scutellum of these immature embryos. The embryogenic callus isolated from the primary explant can be cultured on N6 medium and sub-cultured on this medium every 2 to 3 weeks.

The plasmid, p35S/Ac (obtained from Dr. Peter Eckes, Hoechst Ag, Frankfurt, Germany) may be used in transformation experiments in order to provide for a selectable marker. This plasmid contains the *Pat* gene (see European Patent Publication 0 242 236) which encodes phosphinothricin acetyl transferase (PAT). The enzyme PAT confers resistance to herbicidal glutamine synthetase inhibitors such as phosphinothricin. The *pat* gene in p35S/Ac is under the control of the 35S promoter from cauliflower mosaic virus (Odell et al. (1985) *Nature* 313:810-812) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*.

The particle bombardment method (Klein et al. (1987) *Nature* 327:70-73) may be used to transfer genes to the callus culture cells. According to this method, gold particles (1 µm in diameter) are coated with DNA using the following technique. Ten µg of plasmid DNAs are added to 50 µL of a suspension of gold particles (60 mg per mL). Calcium chloride (50 µL of a 2.5 M solution) and spermidine free base (20 µL of a 1.0 M solution) are added to the particles. The suspension is vortexed during the addition of these solutions. After 10 minutes, the tubes are briefly centrifuged (5 sec at 15,000 rpm) and the supernatant removed. The particles are resuspended in 200 µL of absolute ethanol, centrifuged again and the supernatant removed. The ethanol rinse is performed again and the particles resuspended in a final volume of 30 µL of ethanol. An aliquot (5 µL) of the DNA-coated gold particles can be placed in the center of a Kapton™ flying disc (Bio-Rad

Labs). The particles are then accelerated into the corn tissue with a Biolistic™ PDS-1000/He (Bio-Rad Instruments, Hercules CA), using a helium pressure of 1000 psi, a gap distance of 0.5 cm and a flying distance of 1.0 cm.

For bombardment, the embryogenic tissue is placed on filter paper over agarose-solidified N6 medium. The tissue is arranged as a thin lawn and covered a circular area of about 5 cm in diameter. The petri dish containing the tissue can be placed in the chamber of the PDS-1000/He approximately 8 cm from the stopping screen. The air in the chamber is then evacuated to a vacuum of 28 inches of Hg. The macrocarrier is accelerated with a helium shock wave using a rupture membrane that bursts when the He pressure in the shock tube reaches 1000 psi.

Seven days after bombardment the tissue can be transferred to N6 medium that contains bialaphos (5 mg per liter) and lacks casein or proline. The tissue continues to grow slowly on this medium. After an additional 2 weeks the tissue can be transferred to fresh N6 medium containing bialaphos. After 6 weeks, areas of about 1 cm in diameter of actively growing callus can be identified on some of the plates containing the bialaphos-supplemented medium. These calli may continue to grow when sub-cultured on the selective medium.

Plants can be regenerated from the transgenic callus by first transferring clusters of tissue to N6 medium supplemented with 0.2 mg per liter of 2,4-D. After two weeks the tissue can be transferred to regeneration medium (Fromm et al. (1990) *Bio/Technology* 8:833-839).

EXAMPLE 5

Expression of Recombinant DNA Constructs in Dicot Cells

A seed-specific expression cassette composed of the promoter and transcription terminator from the gene encoding the β subunit of the seed storage protein phaseolin from the bean *Phaseolus vulgaris* (Doyle et al. (1986) *J. Biol. Chem.* 261:9228-9238) can be used for expression of the instant polypeptides in transformed soybean. The phaseolin cassette includes about 500 nucleotides upstream (5') from the translation initiation codon and about 1650 nucleotides downstream (3') from the translation stop codon of phaseolin. Between the 5' and 3' regions are the unique restriction endonuclease sites NcoI (which includes the ATG translation initiation codon), SmaI, KpnI and XbaI. The entire cassette is flanked by HindIII sites.

The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the expression vector.

Amplification is then performed as described above, and the isolated fragment is inserted into a pUC18 vector carrying the seed expression cassette.

Soybean embryos may then be transformed with the expression vector comprising sequences encoding the instant polypeptides. To induce somatic embryos, cotyledons, 3-5 mm in length dissected from surface sterilized, immature seeds of the soybean cultivar A2872, can be cultured in the light or dark at 26 °C on an appropriate agar medium for 6-10 weeks. Somatic embryos which produce secondary embryos are then excised and placed into a suitable liquid medium. After repeated selection for clusters of somatic embryos which multiplied as early, globular staged embryos, the suspensions are maintained as described below.

Soybean embryogenic suspension cultures can be maintained in 35 mL liquid media on a rotary shaker, 150 rpm, at 26 °C with florescent lights on a 16:8 hour day/night schedule. Cultures are subcultured every two weeks by inoculating approximately 35 mg of tissue into 35 mL of liquid medium.

Soybean embryogenic suspension cultures may then be transformed by the method of particle gun bombardment (Klein et al. (1987) *Nature* (London) 327:70-73, U.S. Patent No. 4,945,050). A DuPont Biolistic™ PDS1000/HE instrument (helium retrofit) can be used for these transformations.

A selectable marker gene which can be used to facilitate soybean transformation is a chimeric gene composed of the 35S promoter from cauliflower mosaic virus (Odell et al. (1985) *Nature* 313:810-812), the hygromycin phosphotransferase gene from plasmid pJR225 (from *E. coli*; Gritz et al. (1983) *Gene* 25:179-188) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*. The seed expression cassette comprising the phaseolin 5' region, the fragment encoding the instant polypeptides and the phaseolin 3' region can be isolated as a restriction fragment. This fragment can then be inserted into a unique restriction site of the vector carrying the marker gene.

To 50 µL of a 60 mg/mL 1 µm gold particle suspension is added (in order): 5 µL DNA (1 µg/µL), 20 µL spermidine (0.1 M), and 50 µL CaCl₂ (2.5 M). The particle preparation is then agitated for three minutes, spun in a microfuge for 10 seconds and the supernatant removed. The DNA-coated particles are then washed once in 400 µL 70% ethanol and resuspended in 40 µL of anhydrous ethanol. The DNA/particle suspension can be sonicated three times for one second each. Five µL of the DNA-coated gold particles are then loaded on each macro carrier disk.

Approximately 300–400 mg of a two-week-old suspension culture is placed in an empty 60x15 mm petri dish and the residual liquid removed from the tissue with a pipette. For each transformation experiment, approximately 5–10 plates of tissue are normally bombarded. Membrane rupture pressure is set at 1100 psi and the chamber is evacuated to a vacuum of 28 inches mercury. The tissue is placed approximately 3.5 inches away from the retaining screen and bombarded three times. Following bombardment, the tissue can be divided in half and placed back into liquid and cultured as described above.

Five to seven days post bombardment, the liquid media may be exchanged with fresh media, and eleven to twelve days post bombardment with fresh media containing 50 mg/mL hygromycin. This selective media can be refreshed weekly. Seven to eight weeks post bombardment, green, transformed tissue may be observed growing from untransformed, necrotic embryogenic clusters. Isolated green tissue is removed and inoculated into individual flasks to generate new, clonally propagated, transformed embryogenic suspension cultures. Each new line may be treated as an independent transformation event. These suspensions can then be subcultured and maintained as clusters of immature embryos or regenerated into whole plants by maturation and germination of individual somatic embryos.

EXAMPLE 6

Expression of Recombinant DNA Constructs in Microbial Cells

The cDNAs encoding the instant polypeptides can be inserted into the T7 *E. coli* expression vector pBT430. This vector is a derivative of pET-3a (Rosenberg et al. (1987) *Gene* 56:125-135) which employs the bacteriophage T7 RNA polymerase/T7 promoter system. Plasmid pBT430 was constructed by first destroying the EcoRI and HindIII sites in pET-3a at their original positions. An oligonucleotide adaptor containing EcoRI and Hind III sites was inserted at the BamHI site of pET-3a. This created pET-3aM with additional unique cloning sites for insertion of genes into the expression vector. Then, the NdeI site at the position of translation initiation was converted to an NcoI site using oligonucleotide-directed mutagenesis. The DNA sequence of pET-3aM in this region, 5'-CATATGG, was converted to 5'-CCCATGG in pBT430.

Plasmid DNA containing a cDNA may be appropriately digested to release a nucleic acid fragment encoding the protein. This fragment may then be purified on a 1% low melting agarose gel. Buffer and agarose contain 10 µg/ml ethidium bromide for visualization of the DNA fragment. The fragment can then be purified from the agarose gel by digestion with GELase™ (Epicentre Technologies, Madison,

WI) according to the manufacturer's instructions, ethanol precipitated, dried and resuspended in 20 μ L of water. Appropriate oligonucleotide adapters may be ligated to the fragment using T4 DNA ligase (New England Biolabs (NEB), Beverly, MA). The fragment containing the ligated adapters can be purified from the excess adapters using low melting agarose as described above. The vector pBT430 is digested, dephosphorylated with alkaline phosphatase (NEB) and deproteinized with phenol/chloroform as described above. The prepared vector pBT430 and fragment can then be ligated at 16 °C for 15 hours followed by transformation into DH5 electrocompetent cells (GIBCO BRL). Transformants can be selected on agar plates containing LB media and 100 μ g/mL ampicillin. Transformants containing the gene encoding the instant polypeptides are then screened for the correct orientation with respect to the T7 promoter by restriction enzyme analysis.

For high level expression, a plasmid clone with the cDNA insert in the correct orientation relative to the T7 promoter can be transformed into *E. coli* strain BL21(DE3) (Studier et al. (1986) *J. Mol. Biol.* 189:113-130). Cultures are grown in LB medium containing ampicillin (100 mg/L) at 25 °C. At an optical density at 600 nm of approximately 1, IPTG (isopropylthio- β -galactoside, the inducer) can be added to a final concentration of 0.4 mM and incubation can be continued for 3 h at 25 °C. Cells are then harvested by centrifugation and re-suspended in 50 μ L of 50 mM Tris-HCl at pH 8.0 containing 0.1 mM DTT and 0.2 mM phenyl methylsulfonyl fluoride. A small amount of 1 mm glass beads can be added and the mixture sonicated 3 times for about 5 seconds each time with a microprobe sonicator. The mixture is centrifuged and the protein concentration of the supernatant determined. One μ g of protein from the soluble fraction of the culture can be separated by SDS-polyacrylamide gel electrophoresis. Gels can be observed for protein bands migrating at the expected molecular weight.

CLAIMS

What is claimed is:

1. An isolated polynucleotide comprising:
 - (a) a nucleotide sequence encoding a polypeptide having dihydroflavonol-4-reductase activity, wherein the polypeptide has an amino acid sequence of at least 80% sequence identity, based on the Clustal V method of alignment, when compared to one of SEQ ID NO:6, 8, or 14, or
 - (b) a complement of the nucleotide sequence, wherein the complement and the nucleotide sequence consist of the same number of nucleotides and are 100% complementary.
2. The polynucleotide of Claim 1, wherein the amino acid sequence of the polypeptide has at least 85% sequence identity, based on the Clustal V method of alignment, when compared to one of SEQ ID NO:4, 6, 8, or 14.
3. The polynucleotide of Claim 1, wherein the amino acid sequence of the polypeptide has at least 90% sequence identity, based on the Clustal V method of alignment, when compared to one of SEQ ID NO:4, 6, 8, or 14.
4. The polynucleotide of Claim 1, wherein the amino acid sequence of the polypeptide has at least 95% sequence identity, based on the Clustal V method of alignment, when compared to one of SEQ ID NO:4, 6, 8, 14, or 16.
5. The polynucleotide of Claim 1, wherein the amino acid sequence of the polypeptide has at least 96% sequence identity, based on the Clustal V method of alignment, when compared to one of SEQ ID NO:4, 6, 8, 12, 14, or 16.
6. The polynucleotide of Claim 1, wherein the amino acid sequence of the polypeptide comprises one of SEQ ID NO:4, 6, 8, 12, 14, or 16.
7. The polynucleotide of Claim 1 wherein the nucleotide sequence comprises one of SEQ ID NO:3, 5, 7, 11, 13 or 15.
8. A vector comprising the polynucleotide of Claim 1.
9. A recombinant DNA construct comprising the polynucleotide of Claim 1 operably linked to at least one regulatory sequence.
10. A method for transforming a cell, comprising transforming a cell with the polynucleotide of Claim 1.
 11. A cell comprising the recombinant DNA construct of Claim 9.
 12. A method for producing a plant comprising transforming a plant cell with the polynucleotide of Claim 1 and regenerating a plant from the transformed plant cell.
13. A plant comprising the recombinant DNA construct of Claim 9.
14. A seed comprising the recombinant DNA construct of Claim 9.

15. An isolated polypeptide having dihydroflavonol-4-reductase activity, wherein the polypeptide has an amino acid sequence of at least 80% sequence identity, based on the Clustal V method of alignment, when compared to one of SEQ ID NO:6, 8, or 14.
- 5 16. The polypeptide of Claim 15, wherein the amino acid sequence of the polypeptide has at least 85% sequence identity, based on the Clustal V method of alignment, when compared to one of SEQ ID NO:4, 6, 8, or 14.
- 10 17. The polypeptide of Claim 15, wherein the amino acid sequence of the polypeptide has at least 90% sequence identity, based on the Clustal V method of alignment, when compared to one of SEQ ID NO:4, 6, 8, or 14.
18. The polypeptide of Claim 15, wherein the amino acid sequence of the polypeptide has at least 95% sequence identity, based on the Clustal V method of alignment, when compared to one of SEQ ID NO:4, 6, 8, 14, or 16.
- 15 19. The polypeptide of Claim 15, wherein the amino acid sequence of the polypeptide has at least 96% sequence identity, based on the Clustal V method of alignment, when compared to one of SEQ ID NO:4, 6, 8, 12, 14, or 16.
20. The polypeptide of Claim 15, wherein the amino acid sequence of the polypeptide comprises one of SEQ ID NO:4, 6, 8, 12, 14, or 16.
- 20 21. A method for isolating a polypeptide having dihydroflavonol-4-reductase activity comprising: (a) transforming a cell with a recombinant DNA construct of the present invention; (b) growing the cell in a culture medium; and (c) isolating the polypeptide from the cell or culture medium.
- 25 22. A method of altering the level of expression of a dihydroflavonol-4-reductase in a host cell comprising: (a) transforming a host cell with the recombinant DNA construct of Claim 8; and (b) growing the transformed host cell under conditions that are suitable for expression of the recombinant DNA construct wherein expression of the recombinant DNA construct results in production of altered levels of the dihydroflavonol-4-reductase in the transformed host cell.

30

TITLE

DIHYDROFLAVONOL-4-REDUCTASE

ABSTRACT OF THE DISCLOSURE

- This invention relates to an isolated nucleic acid fragment encoding a
- 5 dihydroflavonol-4-reductase. The invention also relates to the construction of a recombinant DNA construct encoding all or a portion of dihydroflavonol-4-reductase, in sense or antisense orientation, wherein expression of the recombinant DNA construct results in production of altered levels of the dihydroflavonol-4-reductase in a transformed host cell.

10

Figure 1A

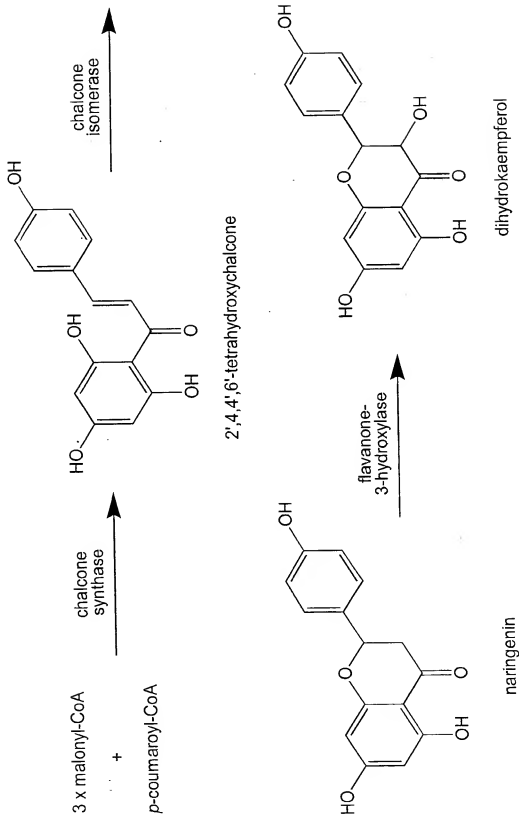


Figure 1B

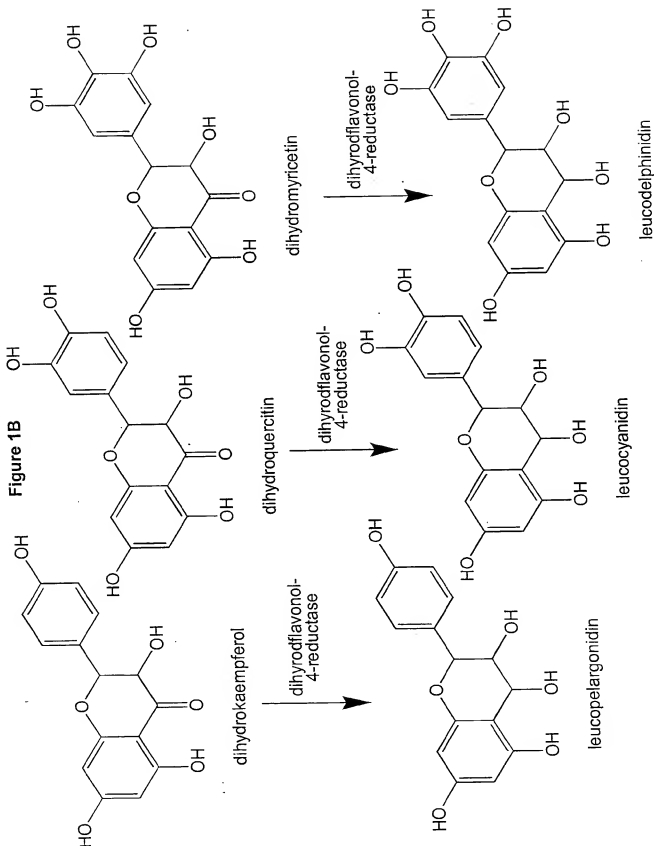


Figure 2A

SEQ ID NO: 4	MSK	10	20	30	40	50	60	55
SEQ ID NO: 6	MSK	10	20	30	40	50	60	55
SEQ ID NO: 8	L	10	20	30	40	50	60	1
SEQ ID NO: 17	MSA	10	20	30	40	50	60	55
SEQ ID NO: 18	MDQTLHTGSKKACVIGTGNLASILIKILLQSGYKVNVTVRDPENEEKIAHLRLQEL	10	20	30	40	50	60	59
SEQ ID NO: 4	SKLSLWKAELETEGSPDEAIKGGCTGVFHLATPVDFKSKDPENEMIKPTTQGVINIMKACL	70	80	90	100	110	120	115
SEQ ID NO: 6	SKLSLWKAELETEGSPDEAIKGGCTGVFHLATPVDFKSKDPENEMIKPTTQGVINIMKACL	70	80	90	100	110	120	115
SEQ ID NO: 8	SKLSLWKAELETEGSPDEAIKGGCTGVFHLATPVDFKSKDPENEMIKPTTQGVINIMKACL	70	80	90	100	110	120	36
SEQ ID NO: 17	SKLSLWKAELETEGSPDEAIKGGCTGVFHLATPVDFKSKDPENEMIKPTTQGVINIMKACL	70	80	90	100	110	120	115
SEQ ID NO: 18	GDLKIFKADLTDEDSPESSFGCEYIFHVAIPINFKSEDPKDMIKPAIQGVINIVLSCL	70	80	90	100	110	120	119
SEQ ID NO: 4	KAKTVRRLLVFTSSAGTTNITEHOKP--IIDETCWTDFEFCRRLLNMTGMWYFVSKTLAEKE	130	140	150	160	170	180	173
SEQ ID NO: 6	KAKTVRRLLVFTSSAGTTNITEHOKP--IIDETCWTDFEFCRRLLNMTGMWYFVSKTLAEKE	130	140	150	160	170	180	173
SEQ ID NO: 8	KAKTVRRLLVFTSSAGTTNITEHOKP--IIDETCWTDFEFCRRLLNMTGMWYFVSKTLAEKE	130	140	150	160	170	180	96
SEQ ID NO: 17	KAKTVRRLLVFTSSAGTTNITEHOKP--IIDETCWTDFEFCRRLLNMTGMWYFVSKTLAEKE	130	140	150	160	170	180	173
SEQ ID NO: 18	KKSKVRVIYITSSAAAVSINNLSGTLVNNENWTDIDFLTEKPFNWCYPISKVLAEEK	130	140	150	160	170	180	179

Figure 2B

SEQ ID NO:4	190	200	210	220	230	240	227
SEQ ID NO:6	---	---	---	---	---	---	227
SEQ ID NO:8	---	---	---	---	---	---	156
SEQ ID NO:17	---	---	---	---	---	---	227
SEQ ID NO:18	---	---	---	---	---	---	239
	+	+	+	+	+	+	
	AWKFAKEHGMDFTAILPALVIGPFLPTMPSSVISALSALSPINGIEAHYSIIKQAQ						
	AWKFAKEHGMDFTAILPALVIGPFLPTMPSSVISALSALSPINGIEAHYSIIKQAQ						
	AWKFAEENHDLITVITPTLTGTGTSVTDTIPSSVGMAASLITGNDFLINALKGMLLSGSI						
	AWKFAKEQGLDFTITPPLVVGPFLLMTPPSSLITALSPTIGNEDHYSIIKQGG						
	AWKFAEENKINLTVITVPALJAGNSLLSDPPSSLSMSFITGKEMHVTGLKMQKLSGSI						
	---	---	---	---	---	---	
	+	+	+	+	+	+	
	---	---	---	---	---	---	
	250	260	270	280	290	300	
	---	---	---	---	---	---	
	+	+	+	+	+	+	
	---	---	---	---	---	---	
	-FVHIEDICLAHFLFEQPKAEGRYICSACDVTIHDIVKLINKEYPKVPTKFNIPDQ						286
	-FVHIEDICLAHFLFEQPKAEGRYICSACDVTIHDIVKLINKEYPKVPTKFNIPDQ						286
	STTHVEDICRAQIFVAEKESASGRVICCANTSVPELAKFLSKRYQKIPTEFDQCSK						216
	-FVHLDLCLAHFLFEPEVEGGRVICSACDATHIDIAKLINOKYPEYKVPKFNIPDQ						286
	SFVHVDDIARAHFLAEKETASGRVICCANTSVPEIADFLIQRYPKNVLSFEFEGLSI						299
	---	---	---	---	---	---	
	+	+	+	+	+	+	
	---	---	---	---	---	---	
	310	320	330	340	350	360	
	---	---	---	---	---	---	
	+	+	+	+	+	+	
	---	---	---	---	---	---	
	LEPVRFSKKITDLGFQFKYSLEDMYTGAIDTCIEKGLPKPAE						337
	LEPVRFSKKITDLGFQFKYSLEDMYTGAIDTCIEKGLPKPAE						337
	-AKLIISSEKLVKEGFSFYXIEEYDQTLFVLSKGL						255
	LELVRFSKKITDLGFQFKYSLEDMYTGAIDTCRDKGLPKPAEGLFKPGETPVNAM						345
	-PKLTSSQLINEGFRFEYGINEMYDQMEYFESKGLI						340
	---	---	---	---	---	---	
	+	+	+	+	+	+	
	---	---	---	---	---	---	

Figure 2C

--
 --
 SEQ ID NO:4 HK
 SEQ ID NO:6 HK
 SEQ ID NO:8 -N
 SEQ ID NO:17 HK
 SEQ ID NO:18 ES

339
 339
 256
 347
 342

Figure 3A

SEQ ID NO:12	M-DGNKGPVVVTGASGFVGSWLYMKLLQGYTVTRATVRDPANVEKNKPLLELPGAKERLS	59
SEQ ID NO:14	MANTSGKVCVTGASGFVASWLVKRLLESGYHVLGTVRDPQGNOKKVAHLWNLAGAKEGLE	60
SEQ ID NO:16	M-DGSGKGPVVVTGASGFVGSWLYMKLLQGYTVTRATVRDPANVEKNKPLLELPGAKERLS	59
SEQ ID NO:19	M-DGNKGPVVVTGASGFVGSWLYMKLLQGYTVTRATVRDPANVEKNKPLLELPGAKERLS	59
SEQ ID NO:20	MWISSKGVKVCVTGASGFVASWLIKRLLLEAGYHVLGTVRDPSPNREKYSVHLWRPLPSAKERLQ	60
SEQ ID NO:12	IWKADLSDEGSFDDAIAGCTGVFHVATPMDFDSDQP--ENEVIKPTVEGMLSTMRACKEAG	118
SEQ ID NO:14	LVRADLLBEGSFDDAVWACGVPFHTASPIITNADS--KEEMLDSAINGTINVIKCKKNP	118
SEQ ID NO:16	IWKADLSDEGSFDDAIAGCTGVFHVATPMDFDSDQP--ENEVIKPTVEGMLSTMRACKEAG	118
SEQ ID NO:19	IWKADLSDEGSFDDAIAGCTGVFHVATPMDFDSDQP--ENEVIKPTVEGMLSTMRACKEAG	118
SEQ ID NO:20	LVRADLMBEGSFDDAVWACGVPFHTASPVIAKSDSDSNCKEEMLVPAINGTILNVLKCKKNP	120
SEQ ID NO:12	TKKRVLTSSSGSTWRLRDEAEFPNVLIDETSWSSVBFCSIQ1--W-YAVAKILAEKSA	174
SEQ ID NO:14	TKKRVLTSSSGSTWRLRDEAEFPNVLIDETSWSSVBFCSIQ1--W-YAVAKILAEKSA	175
SEQ ID NO:16	TKKRVLTSSSGSTWRLRDEAEFPNVLIDETSWSSVBFCSIQ1--W-YAVAKILAEKSA	174
SEQ ID NO:19	TKKRVLTSSSGSTWRLRDEAEFPNVLIDETSWSSVBFCSIQ1--W-YAVAKILAEKSA	174
SEQ ID NO:20	FLKRVLTSSSGSTWRLRDEAEFPNVLIDETSWSSVBFCSIQ1--W-YALAKISAEKAA	176

SEQ	ID	NO:12
SEQ	ID	NO:14
SEQ	ID	NO:16
SEQ	ID	NO:19
SEQ	ID	NO:20

SEQ	ID	NO:12
SEQ	ID	NO:14
SEQ	ID	NO:16
SEQ	ID	NO:19
SEQ	ID	NO:20

SEQ	ID	NO:12
SEQ	ID	NO:14
SEQ	ID	NO:16
SEQ	ID	NO:19
SEQ	ID	NO:20

SEQUENCE LISTING

```

5      <110> Fader, Gary M
        Famodu, Omolayo O.
        Huang, Lisa
        McGonigle, Brian
        Silva, Alexandre da
        Tao, Yong
10
        <120> DIHYDROFLAVONOL-4-REDUCTASE
        <130> BB1521USPRV
15      <140>
        <141>
        <160> 22
20      <170> Microsoft Word Version 7.0A
        <210> 1
        <211> 453
        <212> DNA
25      <213> Glycine max
        <400> 1
        cttgtcttcc agcttgctac acctgtgaac tttgcttctg aagatcctga gaatgacatg 60
        atcaagcctg caatcacagc cgtcttgaat gtgttataaa catgtgtgcg agcaaaaagga 120
30      gtcaaacgag tcattcttgac atcttcagca gctgctgtga ccataaaacca actcaaggga 180
        actgatctgg ttatggatga aagcaactgg actgatgttg agtacttgag cactgcaaaag 240
        ccaccactt ggggttatcc tgcctccaaa gcactagcag agaaggcagc atggaaattt 300
        gctgaagaaa atcacattga tctcatcact gtgataccta ctctcacaac tggtccttct 360
        gtccactcaa gacatccaa caagtgttgg catggcaacg tccctcaaaa caggcaatga 420
35      tttcctcaaa aacgccctga aagggtatgc aaa 453
        <210> 2
        <211> 139
        <212> PRT
40      <213> Glycine max
        <400> 2
        Leu Val Phe Gln Leu Ala Thr Pro Val Asn Phe Ala Ser Glu Asp Pro
45      1 5 10 15
        Glu Asn Asp Met Ile Lys Pro Ala Ile Thr Gly Val Leu Asn Val Leu
        20 25 30
50      Lys Ala Cys Val Arg Ala Lys Gly Val Lys Arg Val Ile Leu Thr Ser
        35 40 45
        Ser Ala Ala Ala Val Thr Ile Asn Gln Leu Lys Gly Thr Asp Leu Val
        50 55 60
55      Met Asp Glu Ser Asn Trp Thr Asp Val Glu Tyr Leu Ser Thr Ala Lys
        65 70 75 80
        Pro Pro Thr Trp Gly Tyr Pro Ala Ser Lys Ala Leu Ala Glu Lys Ala

```

Ala Trp Lys Phe Ala Glu Glu Asn His Ile Asp Leu Ile Thr Val Ile
100 105 110

5

Pro Thr Leu Thr Thr Gly Pro Ser Val Thr Thr Arg His Pro Asn Lys
115 120 125

10

Cys Trp His Gly Asn Val Pro Gln Asn Arg Gln
130 135

<210> 3

<211> 1234

15

<212> DNA

<213> Glycine max

<400> 3

20

gcacgagatt ttatttttctt ttctttcttt ggaagataaa gaatgggttc taagtccgaa 60

accggtttgc ttactggggc ttctgggttac atcggatcat ggcttgctcat gagactcatc 120

gagcgtggct ataccggttc agccaccgta ctcgacccag ctgatattgag ggagggtgaag 180

catttgctcgt atctgccagg tgcagagagc aagctgtctc tgtggaaggc agaacttaca 240

gaagagggaa gctttgatga agccattaaa ggggtgcacag gtgttttcca cttggccacc 300

25

cccgttgact ttaagtccaa agaccagag aatgaaatga taaagcctac aattcaagga 360

gtactaaaca tcatgaaagc atgcctgaag gcaaaaactg tccgaaggct agtattcacg 420

tcctcagcgc gaactaccaa cattactgag caccaaaagc ctatcattga cgaaacctgc 480

tggactgatg ttgagttctg ccggagatta aatatgactg gttggatgta ttctgttctc 540

aaaacacttg cggagaaga agcttggaat tttgcgaagc agcacggcat ggacttcatc 600

30

gctatccttc cagctcttgt cattggtccc ttctactgc caacaatgcc ttctagcgtg 660

atcagtgctc ttccacctat taacggaatt gaggcacatt attcaatcat aaagcaagct 720

caattcgtcc acatagaaga tatctgtctt gctcacatat ttctgtttga acagccaaaa 780

gcagaaggga ggtatatatg cagtgcatgt gacgttacta tccatgacat tgtaaaatta 840

attaacgaaa aataccaga gtacaagggt cccaccaagt ttcagaacat tccagatcaa 900

35

ttggagcccg tgagatttcc ttccaagaaa atcacagact tgggattcca atttaaatatc 960

agcttagagg atatgtacac tggagcaatt gatacatgca tagagaaagg gctttctctc 1020

aaacctgcag aaattccagc gaatggcatc gagcataaat aaatataggt ttctcatatc 1080

ttgctcgtg gatggctatg aatgttgctt ttctgtttca gtttctttaa tgaatgtttc 1140

gttttgtgaa ttctgatgca aaattgtaag tgggtttgtaa gaccaaaatta gttatctaca 1200

40

aattgtttaa tattatcaca aaaaaaaaaa aaaa 1234

<210> 4

<211> 339

45

<212> PRT

<213> Glycine max

<400> 4

Met Gly Ser Lys Ser Glu Thr Val Cys Val Thr Gly Ala Ser Gly Tyr

50

1 5 10 15

Ile Gly Ser Trp Leu Val Met Arg Leu Ile Glu Arg Gly Tyr Thr Val

20 25 30

Arg Ala Thr Val Leu Asp Pro Ala Asp Met Arg Glu Val Lys His Leu

55

35 40 45

Leu Asp Leu Pro Gly Ala Glu Ser Lys Leu Ser Leu Trp Lys Ala Glu

50 55 60

60

Leu Thr Glu Glu Gly Ser Phe Asp Glu Ala Ile Lys Gly Cys Thr Gly

	65	70	75	80
	Val Phe His Leu Ala Thr Pro Val Asp Phe Lys Ser Lys Asp Pro Glu			
		85	90	95
5	Asn Glu Met Ile Lys Pro Thr Ile Gln Gly Val Leu Asn Ile Met Lys			
		100	105	110
10	Ala Cys Leu Lys Ala Lys Thr Val Arg Arg Leu Val Phe Thr Ser Ser			
		115	120	125
	Ala Gly Thr Thr Asn Ile Thr Glu His Gln Lys Pro Ile Ile Asp Glu			
		130	135	140
15	Thr Cys Trp Thr Asp Val Glu Phe Cys Arg Arg Leu Asn Met Thr Gly			
		145	150	155
	Trp Met Tyr Phe Val Ser Lys Thr Leu Ala Glu Lys Glu Ala Trp Lys			
		165	170	175
20	Phe Ala Lys Glu His Gly Met Asp Phe Ile Ala Ile Leu Pro Ala Leu			
		180	185	190
25	Val Ile Gly Pro Phe Leu Leu Pro Thr Met Pro Ser Ser Val Ile Ser			
		195	200	205
	Ala Leu Ser Pro Ile Asn Gly Ile Glu Ala His Tyr Ser Ile Ile Lys			
		210	215	220
30	Gln Ala Gln Phe Val His Ile Glu Asp Ile Cys Leu Ala His Ile Phe			
		225	230	235
	Leu Phe Glu Gln Pro Lys Ala Glu Gly Arg Tyr Ile Cys Ser Ala Cys			
		245	250	255
35	Asp Val Thr Ile His Asp Ile Val Lys Leu Ile Asn Glu Lys Tyr Pro			
		260	265	270
40	Glu Tyr Lys Val Pro Thr Lys Phe Gln Asn Ile Pro Asp Gln Leu Glu			
		275	280	285
	Pro Val Arg Phe Ser Ser Lys Lys Ile Thr Asp Leu Gly Phe Gln Phe			
		290	295	300
45	Lys Tyr Ser Leu Glu Asp Met Tyr Thr Gly Ala Ile Asp Thr Cys Ile			
		305	310	315
	Glu Lys Gly Leu Leu Pro Lys Pro Ala Glu Ile Pro Ala Asn Gly Ile			
		325	330	335
50	Glu His Lys			
55	<210> 5			
	<211> 1389			
	<212> DNA			
	<213> Glycine max			
60	<400> 5			

gcaccagaaa attatctcca tcattttcat cgtgccggtc ctcaactctc getccaacaa 60
 ccaatgccta caactacaag gctacaacaa agcatctata aaagataaag aatgggttct 120
 aagtcggaaa ccgtttcggt tactgggggt tctggttaca tcggatcatg gcttgtcatg 180
 agactcatcg agcgtggcta taccgttcga gccaccgtac tcgacccagc tgatatgagg 240
 gaggtgaagc attttgctgga tctgccagggt gcagagagca agctgtctct gtggaaggca 300
 gaacttacag aagaggggaag ctttgatgaa gccattaaag ggtgcacagg tgttttccac 360
 ttggccaccc ccgttgactt taagtccaaa gacccagaga atgaaatgat aaagcctaca 420
 attcaaggag tactaaacat catgaaagca tgcctgaaag caaaaactgt ccgaaggcta 480
 gtattcacgt cctcagccgg aactaccaac attactgagc accaaaagcc tatcattgac 540
 gaaacctgct ggactgatgt tgagttctgc cggagattaa atatgactgg ttggatgtat 600
 ttcgtttcta aaacacttgc ggagaagaa gcttggaat ttgcgaaaga gcacggcatg 660
 gacttcatcg ctatccttcc agctcttctc attggtccct ttctactgcc aacaatacct 720
 tctagcgtga tcaagtgtct ttcacctatt aacggaattg aggcacatta ttcaatcata 780
 aagcaagctc aattcgtcca catagaagat atctgtcttg ctccatatt tctgtttgaa 840
 cagccaaaag cagaagggag gtatatatgc agtgcattgt acgttactat ccatgacatt 900
 gtaaaattaa ttaacgaaaa ataccagag tacaaggctc ccaccaagtt tcagaacatt 960
 ccagatcaat tggagcccggt gagattttct tccaagaaaa tcacagactt gggattccaa 1020
 tttaaataca gcttagagga tatgtacact ggagcaattg atacatgcat agagaagggg 1080
 tcttctccta aacctgcaga aattccagcg aatggcatcg agcataaata aatataggtt 1140
 ttcatacttt tgcctcgggtg atggctatga atgttgcctt tctgtttcag tttctttaa 1200
 gatgtttccg ttttgatgaa tctgtagtaa aattgtaagt ggtttgtaag accaaaattag 1260
 ttatctacaa attgtttaat attatcaca gacatttttt acggtaaagc aatatctcag 1320
 agctcgaggc aatttaaaat atataataa tgcgcgcatt gaattctcat taaaaaaa 1380
 aaaaaaaa 1389

<210> 6
 <211> 339
 <212> PRT
 <213> Glycine max

<400> 6
 Met Gly Ser Lys Ser Glu Thr Val Cys Val Thr Gly Ala Ser Gly Tyr
 1 5 10 15
 Ile Gly Ser Trp Leu Val Met Arg Leu Ile Glu Arg Gly Tyr Thr Val
 20 25 30
 Arg Ala Thr Val Leu Asp Pro Ala Asp Met Arg Glu Val Lys His Leu
 35 40 45
 Leu Asp Leu Pro Gly Ala Glu Ser Lys Leu Ser Leu Trp Lys Ala Glu
 50 55 60
 Leu Thr Glu Glu Gly Ser Phe Asp Glu Ala Ile Lys Gly Cys Thr Gly
 65 70 75 80
 Val Phe His Leu Ala Thr Pro Val Asp Phe Lys Ser Lys Asp Pro Glu
 85 90 95
 Asn Glu Met Ile Lys Pro Thr Ile Gln Gly Val Leu Asn Ile Met Lys
 100 105 110
 Ala Cys Leu Lys Ala Lys Thr Val Arg Arg Leu Val Phe Thr Ser Ser
 115 120 125
 Ala Gly Thr Thr Asn Ile Thr Glu His Gln Lys Pro Ile Ile Asp Glu
 130 135 140
 Thr Cys Trp Thr Asp Val Glu Phe Cys Arg Arg Leu Asn Met Thr Gly

	145		150		155		160
	Trp Met Tyr Phe Val Ser Lys Thr Leu Ala Glu Lys Glu Ala Trp Lys						
		165			170		175
5	Phe Ala Lys Glu His Gly Met Asp Phe Ile Ala Ile Leu Pro Ala Leu			185		190	
		180					
10	Val Ile Gly Pro Phe Leu Leu Pro Thr Ile Pro Ser Ser Val Ile Ser			200		205	
		195					
	Ala Leu Ser Pro Ile Asn Gly Ile Glu Ala His Tyr Ser Ile Ile Lys			215		220	
		210					
15	Gln Ala Gln Phe Val His Ile Glu Asp Ile Cys Leu Ala His Ile Phe			230		235	240
	Leu Phe Glu Gln Pro Lys Ala Glu Gly Arg Tyr Ile Cys Ser Ala Cys			245		250	255
20	Asp Val Thr Ile His Asp Ile Val Lys Leu Ile Asn Glu Lys Tyr Pro			260		265	270
25	Glu Tyr Lys Val Pro Thr Lys Phe Gln Asn Ile Pro Asp Gln Leu Glu			275		280	285
	Pro Val Arg Phe Ser Ser Lys Lys Ile Thr Asp Leu Gly Phe Gln Phe			290		295	300
30	Lys Tyr Ser Leu Glu Asp Met Tyr Thr Gly Ala Ile Asp Thr Cys Ile			305		310	315
	Glu Lys Gly Leu Leu Pro Lys Pro Ala Glu Ile Pro Ala Asn Gly Ile			320		325	330
35	Glu His Lys						
40	<210> 7						
	<211> 952						
	<212> DNA						
	<213> Glycine max						
45	<400> 7						
	cttgtctctcc agcctgtgtac acctgtgaac ttgtctctcg aagatcctga gaatgacatg 60						
	atcaagcctg caatcacagg cgtcttgaat gtgttataaag catgtgtgcg agcaaaaagga 120						
	gtcaaacagag tcatcttgac atcttcagca gctgctgtga ccataaacca actcaaggga 180						
50	actgatctgg ttatggatga aagcaactgg actgatgttg agtacttgag cactgcaaaag 240						
	ccaccacttt ggggttatcc tgctcccaaa gcactagcag agaaggcagc atggaaaattt 300						
	gtgaagaaaa atcacattga tctcatcact gtgataccta ctctcacaaac tggctccttct 360						
	gtcactacag acatcccatc aagtgtttggc atggcagcgt ccctcataac agggcaatgat 420						
	ttctcataaa acgctctgaa aggtatgcag ttgctatcag gttcaatatc catcactcat 480						
55	gtggaggata ttgtccgagc acaaatattt gtggcggaga aagaatcggc ttctgtgtcga 540						
	tacatttgcg gtgctcacaa tactagtgtt cctgagcttg caaagtttct cagcaaacga 600						
	taccttcaat ataaaattcc aactgaattc gatgattgtc cctcaaaggc aaagttaata 660						
	atctcttccg aaaaagcttg caaagaagga ttacagttta agtattggaat tgaagaaatt 720						
	tatgatcaga cctctggagta cttaaagagc aaaaggggctc tgaataactg aattttgaaa 780						
	attctaatgc tctagcaaac tctccacttg ttatgtagta gcactcttgc aataattaat 840						
60	caaggggaat agcaaacatc tctgttaatc cactacttta ttaccacctc cttaaattac 900						

tgattatgat attccccact ttgttaaaaa aaaaaaaaaa aaaaaaaaaa aa

952

5 <210> 8
 <211> 256
 <212> PRT
 <213> Glycine max

 <400> 8
 10 Leu Val Phe Gln Leu Ala Thr Pro Val Asn Phe Ala Ser Glu Asp Pro
 1 5 10 15
 Glu Asn Asp Met Ile Lys Pro Ala Ile Thr Gly Val Leu Asn Val Leu
 20 25 30
 15 Lys Ala Cys Val Arg Ala Lys Gly Val Lys Arg Val Ile Leu Thr Ser
 35 40 45
 Ser Ala Ala Ala Val Thr Ile Asn Gln Leu Lys Gly Thr Asp Leu Val
 50 55 60
 Met Asp Glu Ser Asn Trp Thr Asp Val Glu Tyr Leu Ser Thr Ala Lys
 65 70 75 80
 25 Pro Pro Thr Trp Gly Tyr Pro Ala Ser Lys Ala Leu Ala Glu Lys Ala
 85 90 95
 Ala Trp Lys Phe Ala Glu Glu Asn His Ile Asp Leu Ile Thr Val Ile
 100 105 110
 30 Pro Thr Leu Thr Thr Gly Pro Ser Val Thr Thr Asp Ile Pro Ser Ser
 115 120 125
 Val Gly Met Ala Ala Ser Leu Ile Thr Gly Asn Asp Phe Leu Ile Asn
 130 135 140
 Ala Leu Lys Gly Met Gln Leu Leu Ser Gly Ser Ile Ser Ile Thr His
 145 150 155 160
 40 Val Glu Asp Ile Cys Arg Ala Gln Ile Phe Val Ala Glu Lys Glu Ser
 165 170 175
 Ala Ser Gly Arg Tyr Ile Cys Cys Ala His Asn Thr Ser Val Pro Glu
 180 185 190
 45 Leu Ala Lys Phe Leu Ser Lys Arg Tyr Pro Gln Tyr Lys Ile Pro Thr
 195 200 205
 Glu Phe Asp Asp Cys Pro Ser Lys Ala Lys Leu Ile Ser Ser Glu
 210 215 220
 Lys Leu Val Lys Glu Gly Phe Ser Phe Lys Tyr Gly Ile Glu Glu Ile
 225 230 235 240
 55 Tyr Asp Gln Thr Leu Glu Tyr Leu Lys Ser Lys Gly Ala Leu Asn Asn
 245 250 255

 <210> 9
 60 <211> 546


```

<212> DNA
<213> Triticum aestivum

5  <220>
    <221> unsure
    <222> (534)..(535)..(536)..(537)..(538)
    <223> n = a, c, g or t

10 <220>
    <221> unsure
    <222> (542)..(543)..(544)
    <223> n = a, c, g or t

15 <400> 9
    gcacgaggga gagaatggcg aatacttcca agggcaaggt gtgtgtcact ggagcctctg 60
    gctttgttgc ctcttggcct gtcaagcgac ttctcagtc aggttatcat gtcttaggga 120
    cagtcagaga cccaggcaat cagaagaagg tagcacacct ctggaactta gcagggtgcta 180
    agggggggct ggagcttgct agggctgacc tcttgggaaga agggagcttc gacgatgccg 240
    tgatggcctg tgagggcgct ttccacaccg catcgccctat catcactaac gctgattcca 300
    aggaagaaat gcttgattcg gcgataaacg gcactctaaa cgtgctgaga tcttgcaaga 360
    agaattccatt cctcaaaagg gttgtcctca cgtcttctgc gtcgacctat aggcctgagag 420
    atgaagctga attccctccc aacgtgctgc tggatgaaac atcatggagc tccgtggagt 480
    tctgcgaaag catccagata tggtaacgctg ttgcgaagat cctcgctgag aaannnnncc 540
    tnnnag                                     546

25 <210> 10
    <211> 177
    <212> PRT
30 <213> Triticum aestivum

    <220>
    <221> UNSURE
    <222> (174)..(175)
35 <223> Xaa = ANY AMINO ACID

    <220>
    <221> UNSURE
    <222> (177)
40 <223> Xaa = ANY AMINO ACID

    <400> 10
    Met Ala Asn Thr Ser Lys Gly Lys Val Cys Val Thr Gly Ala Ser Gly
      1          5          10          15

45 Phe Val Ala Ser Trp Leu Val Lys Arg Leu Leu Glu Ser Gly Tyr His
      20          25          30

    Val Leu Gly Thr Val Arg Asp Pro Gly Asn Gln Lys Lys Val Ala His
      35          40          45

    Leu Trp Asn Leu Ala Gly Ala Lys Glu Gly Leu Glu Leu Val Arg Ala
      50          55          60

55 Asp Leu Leu Glu Glu Gly Ser Phe Asp Asp Ala Val Met Ala Cys Glu
      65          70          75          80

    Gly Val Phe His Thr Ala Ser Pro Ile Ile Thr Asn Ala Asp Ser Lys
      85          90          95

60

```

Glu Glu Met Leu Asp Ser Ala Ile Asn Gly Thr Leu Asn Val Leu Arg
 100 105 110
 5 Ser Cys Lys Lys Asn Pro Phe Leu Lys Arg Val Val Leu Thr Ser Ser
 115 120 125
 Ser Ser Thr Met Arg Leu Arg Asp Glu Ala Glu Phe Pro Pro Asn Val
 130 135 140
 10 Leu Leu Asp Glu Thr Ser Trp Ser Ser Val Glu Phe Cys Glu Ser Ile
 145 150 155 160
 Gln Ile Trp Tyr Ala Val Ala Lys Ile Leu Ala Glu Lys Xaa Xaa Leu
 165 170 175
 15 Xaa
 20 <210> 11
 <211> 1291
 <212> DNA
 <213> Triticum aestivum
 25 <400> 11
 gaaagctcaa gaaggaagtc gatagagaga tgaaggagga gacaatggac gggaacaaaag 60
 gaccgggtggt ggtgaccgga gcgtcgggtt tcgtaggatc atggctcgtc atgaagctcc 120
 tccagggtcgg gtacaccgtc cgggccaccg tgcgcgaccc ggccaacggt gagaagaaca 180
 30 agccattgct ggagcttccc ggagccaagg agcggctgtc catctggaag gccgacctga 240
 gcgaggaagg cagctctcac gacgccatcg ccggctgcac cggcgtcttc cactctcgca 300
 gcgccatgga cttcgactcc caagaccocg aaaacgaggt gatcaagccg acggtggaag 360
 ggatgctgag catcatgagg gccctgcaagg aggcctggcac cgtgaaacgc atagtcttca 420
 cotctcccg cggcagcgtc aacatcgagg agcggcagcg gccagcctac gaccaggaca 480
 actggagcga catcgacttc tgcgcgcgcg tcaagatgac aggatggatg tacttctgtg 540
 35 ccaagtcctt ggcagagaag gccgccatgg agtacgccag cgagaacggc ctggacttca 600
 tcagcatcat ccccacgctc gtctctggcc cgttctctag cgcggccatg ccgcccagcc 660
 tgcacccgc cctcgcgtc atcacgggga acgaggccca ctactcgatc ctgaagcagg 720
 tgcagctggt ccacctggat gacctctgcg acgccatgac ctctctcttc gagcaccocg 780
 aggcacaagg ccgtacatc tgctctctcc acgacgccac catccacggc ctgccaggga 840
 40 tgctccggga caggttcccc gactacagca tcccacagaa ttctcgaggt gtcgacgacg 900
 acctccagcc catccacttc tcttccaaga agctctctga ccacggcttc agcttccggg 960
 acaccgcgga ggacatgttc gacgcccgca tccgcacctg cagggagaag ggcctcattc 1020
 cgctcggaga cgcccgcgcg cctgcagccg cggcgcaagct gggagctctt gtcgccccgg 1080
 aaggccaagg cattgtgtga gaaacataa aagccagcgc tgctgcgtga atactattct 1140
 45 tgtgtctgca atttgcattg gcagagccct gtaactagtg ggatcatcat gactatggag 1200
 tgcattcatt tttttccacc tcggtagtat tatgaaataa aattgaaat aacgcataaa 1260
 ctttatcagt attaaaaaaa aaaaaaaaaa a
 1291
 50 <210> 12
 <211> 354
 <212> PRT
 <213> Triticum aestivum
 55 <400> 12
 Met Asp Gly Asn Lys Gly Pro Val Val Val Thr Gly Ala Ser Gly Phe
 1 5 10 15
 Val Gly Ser Trp Leu Val Met Lys Leu Leu Gln Val Gly Tyr Thr Val
 20 25 30

Arg Ala Thr Val Arg Asp Pro Ala Asn Val Glu Lys Asn Lys Pro Leu
 35 40 45
 5 Leu Glu Leu Pro Gly Ala Lys Glu Arg Leu Ser Ile Trp Lys Ala Asp
 50 55 60
 Leu Ser Glu Glu Gly Ser Phe Asp Asp Ala Ile Ala Gly Cys Thr Gly
 65 70 75 80
 10 Val Phe His Val Ala Thr Pro Met Asp Phe Asp Ser Gln Asp Pro Glu
 85 90 95
 Asn Glu Val Ile Lys Pro Thr Val Glu Gly Met Leu Ser Ile Met Arg
 100 105 110
 Ala Cys Lys Glu Ala Gly Thr Val Lys Arg Ile Val Phe Thr Ser Ser
 115 120 125
 20 Ala Gly Ser Val Asn Ile Glu Glu Arg Gln Arg Pro Ala Tyr Asp Gln
 130 135 140
 Asp Asn Trp Ser Asp Ile Asp Phe Cys Arg Arg Val Lys Met Thr Gly
 145 150 155 160
 25 Trp Met Tyr Phe Val Ser Lys Ser Leu Ala Glu Lys Ala Ala Met Glu
 165 170 175
 Tyr Ala Ser Glu Asn Gly Leu Asp Phe Ile Ser Ile Ile Pro Thr Leu
 180 185 190
 30 Val Val Gly Pro Phe Leu Ser Ala Gly Met Pro Pro Ser Leu Val Thr
 195 200 205
 Ala Leu Ala Leu Ile Thr Gly Asn Glu Ala His Tyr Ser Ile Leu Lys
 210 215 220
 Gln Val Gln Leu Val His Leu Asp Asp Leu Cys Asp Ala Met Thr Phe
 225 230 235 240
 40 Leu Phe Glu His Pro Glu Ala Asn Gly Arg Tyr Ile Cys Ser Ser His
 245 250 255
 Asp Ala Thr Ile His Gly Leu Ala Arg Met Leu Arg Asp Arg Phe Pro
 260 265 270
 45 Glu Tyr Ser Ile Pro Gln Lys Phe Ala Gly Val Asp Asp Asp Leu Gln
 275 280 285
 50 Pro Ile His Phe Ser Ser Lys Lys Leu Leu Asp His Gly Phe Ser Phe
 290 295 300
 Arg Tyr Thr Ala Glu Asp Met Phe Asp Ala Ala Ile Arg Thr Cys Arg
 305 310 315 320
 55 Glu Lys Gly Leu Ile Pro Leu Gly Asp Ala Pro Pro Pro Ala Ala Gly
 325 330 335
 60 Gly Lys Leu Gly Ala Leu Ala Ala Gly Glu Gly Gln Ala Ile Gly Ala
 340 345 350

Glu Thr

5 <210> 13
 <211> 1222
 <212> DNA
 <213> Triticum aestivum

10 <400> 13
 gcacgaggca gagaatggcg aatacttcca agggcaaggt gtgtgtcact ggagcctctg 60
 gctttgttgc ctctttggctt gtcaagcgac ttctcgagtc aggttatcat gttctagggg 120
 cagtcagaga cccaggcaat cagaagaagg tagcacacct ctggaactta gcagggtgcta 180
 15 agggaggggct ggagcttgtc aggggtgacc tcttggaaga agggagcttc gacgatgccg 240
 tgatggcctg tgagggcgctc ttccacaccg catcgccctat catcactaac gctgattcca 300
 aggaagaaat gcttgattcg gcgataaacg gcactctaaa cgtgctgaga tcttgcaaga 360
 agaatccatt cctcaaaaagg gttgtctca cgtcttcgtc gtcgacctg aggtcgagag 420
 atgaagctga attccctccc aacgtgctgc tggatgaaac atcatggagc tccgtggagt 480
 20 tctgcgaaaag catccagata tgggtacgctg ttgcgaagat cctcgctgag aaatcagcct 540
 gggagttcgc caaggagaac aacatcgacc tcgtggccgt tcttccgacg ttcatcatcg 600
 gccctaacct ctccccctgtg ttaggcccca ccgcctcaga tgtccttggc ttgtttaaaag 660
 gggagacgga gaagttcacc atcttcggga ggatggggta cgtccacatc gacgacgtcg 720
 cgagctgcca catcctggctc tacgaaaccg ccgacgcca ggaggaggtac atctgcaact 780
 25 cggcggttct ggatagcaac gagctggctg ccttgcctgc gaaacgggtc ccgtcgcttc 840
 ccattcccga gagtttaccg aacatttcag gggagcgac gtacggctac aacacgtcaa 900
 agatcccgaa gctggggctt gagttcagag gcgtggagga gatgttcgac gactcggtgg 960
 agtcgctcaa gccgcatggc tatctgcgcg agggcgcccg gtgatcgga tgagcggaca 1020
 ggaatcgatcc gccggagggg agaatacgag gccctgggtg ctgctctaga agcatttcag 1080
 30 ttattactaa cacatgagaa gaatcgcttc cctttctgtt ctctatttct gcttgggtg 1140
 ttcccttgcg actctgcaga cagtcatgaa cgacaaggaa atgaaatgta ctccatgttg 1200
 cattaataaaa aaaaaaaaaa aa 1222

35 <210> 14
 <211> 329
 <212> PRT
 <213> Triticum aestivum

40 <400> 14
 Met Ala Asn Thr Ser Lys Gly Lys Val Cys Val Thr Gly Ala Ser Gly
 1 5 10 15

45 Phe Val Ala Ser Trp Leu Val Lys Arg Leu Leu Glu Ser Gly Tyr His
 20 25 30

Val Leu Gly Thr Val Arg Asp Pro Gly Asn Gln Lys Lys Val Ala His
 35 40 45

50 Leu Trp Asn Leu Ala Gly Ala Lys Glu Gly Leu Glu Leu Val Arg Ala
 50 55 60

Asp Leu Leu Glu Glu Gly Ser Phe Asp Asp Ala Val Met Ala Cys Glu
 65 70 75 80

55 Gly Val Phe His Thr Ala Ser Pro Ile Ile Thr Asn Ala Asp Ser Lys
 85 90 95

60 Glu Glu Met Leu Asp Ser Ala Ile Asn Gly Thr Leu Asn Val Leu Arg
 100 105 110

Ser Cys Lys Lys Asn Pro Phe Leu Lys Arg Val Val Leu Thr Ser Ser
 115 120 125
 5 Ser Ser Thr Met Arg Leu Arg Asp Glu Ala Glu Phe Pro Pro Asn Val
 130 135 140
 Leu Leu Asp Glu Thr Ser Trp Ser Ser Val Glu Phe Cys Glu Ser Ile
 145 150 155 160
 10 Gln Ile Trp Tyr Ala Val Ala Lys Ile Leu Ala Glu Lys Ser Ala Trp
 165 170 175
 Glu Phe Ala Lys Glu Asn Asn Ile Asp Leu Val Ala Val Leu Pro Thr
 180 185 190
 15 Phe Ile Ile Gly Pro Asn Leu Ser Pro Val Leu Gly Pro Thr Ala Ser
 195 200 205
 20 Asp Val Leu Gly Leu Phe Lys Gly Glu Thr Glu Lys Phe Thr Ile Phe
 210 215 220
 Gly Arg Met Gly Tyr Val His Ile Asp Asp Val Ala Ser Cys His Ile
 225 230 235 240
 25 Leu Val Tyr Glu Thr Ala Asp Ala Lys Gly Arg Tyr Ile Cys Asn Ser
 245 250 255
 30 Ala Val Leu Asp Ser Asn Glu Leu Val Ala Leu Leu Ala Lys Arg Phe
 260 265 270
 Pro Ser Phe Pro Ile Pro Lys Ser Leu Pro Asn Ile Tyr Gly Glu Gln
 275 280 285
 35 Thr Tyr Gly Tyr Asn Thr Ser Lys Ile Arg Lys Leu Gly Leu Glu Phe
 290 295 300
 Arg Gly Val Glu Glu Met Phe Asp Asp Ser Val Glu Ser Leu Lys Ala
 305 310 315 320
 40 His Gly Tyr Leu Arg Glu Gly Ala Ala
 325
 45 <210> 15
 <211> 1350
 <212> DNA
 <213> Triticum aestivum
 50 <400> 15
 gcacgagagc cattgaaaga aagctcaaga aggaagtcga tagagagatg aaggggagaga 60
 caatggacgg gagcaaaagg cgggtgggtg tgactggagc gtcggggtttc gtaggggtcgt 120
 ggctcgtcat gaagctcctc caggccgggt acacgcgtcc ggccaccgtg cgcgaccccg 180
 ccaacgttga gaagaacaag cggttgctgg agcttccccg agccaaggag cggctgtcca 240
 55 tctggaaggc cgactgcagc gacgaaggca gcttcgacga cgccatcgcc ggctgcaccg 300
 gcgtcttcca cgctcgccacg cccatggact tcgaactccaa agatccccgag aacgaggtga 360
 tcaaaaccac ggtggaaggg atgctgagca tcatgagggc ctgcaaggag gctggcaccg 420
 tgaagcgcat cgtcttcacc tcctccgcgc gcagcgtcaa catcgaggag cggcagcgcc 480
 cagcctacga ccaggacaac tggagcgaca tcgactctcg ccgcccgctc aagatgacag 540
 60 gatggatgta cttcgtgtcc aagtcctcgc cagagaaggc cgccatggag tacgccagcg 600

Ala Leu Ala Leu Ile Thr Gly Asn Glu Ala His Tyr Ser Ile Leu Lys
 210 215 220
 5 Gln Val Gln Leu Val His Leu Asp Asp Leu Cys Asp Ala Met Thr Phe
 225 230 235 240
 Leu Phe Glu His Pro Glu Ala Asn Gly Arg Tyr Ile Cys Ser Ser His
 245 250 255
 10 Asp Ala Thr Ile His Gly Leu Ala Arg Met Leu Arg Asp Arg Phe Pro
 260 265 270
 Glu Tyr Ser Ile Pro His Lys Phe Ala Gly Val Asp Asp Leu Gln
 275 280 285
 15 Pro Ile His Phe Ser Ser Lys Lys Leu Leu Asp His Gly Phe Ser Phe
 290 295 300
 20 Arg Tyr Thr Ala Glu Asp Met Phe Asp Ala Ala Ile Arg Thr Cys Arg
 305 310 315 320
 Glu Lys Gly Leu Ile Pro Leu Gly Asp Ala Pro Ala Pro Ala Ala
 325 330 335
 25 Gly Lys Leu Gly Ala Leu Ala Ala Gly Lys Gly Gln Ala Ile Gly Ala
 340 345 350
 Glu Thr
 30
 <210> 17
 <211> 347
 <212> PRT
 35 <213> Glycine max
 <400> 17
 Met Gly Ser Ala Ser Glu Ser Val Cys Val Thr Gly Ala Ser Gly Phe
 1 5 10 15
 40 Ile Gly Ser Trp Leu Val Met Arg Leu Ile Glu Arg Gly Tyr Thr Val
 20 25 30
 45 Arg Ala Thr Val Arg Asp Pro Val Asn Met Lys Lys Val Lys His Leu
 35 40 45
 Val Glu Leu Pro Gly Ala Lys Ser Lys Leu Ser Leu Trp Lys Ala Asp
 50 55 60
 50 Leu Ala Glu Glu Gly Ser Phe Asp Glu Ala Ile Lys Gly Cys Thr Gly
 65 70 75 80
 Val Phe His Val Ala Thr Pro Met Asp Phe Glu Ser Lys Asp Pro Glu
 85 90 95
 55 Asn Glu Val Ile Lys Pro Thr Ile Asn Gly Val Leu Asp Ile Met Lys
 100 105 110
 60 Ala Cys Leu Lys Ala Lys Thr Val Arg Arg Leu Ile Phe Thr Ser Ser
 115 120 125

Ala Gly Thr Leu Asn Val Ile Glu Arg Gln Lys Pro Val Phe Asp Asp
 130 135 140
 5 Thr Cys Trp Ser Asp Val Glu Phe Cys Arg Arg Val Lys Met Thr Gly
 145 150 155 160
 Trp Met Tyr Phe Val Ser Lys Thr Leu Ala Glu Lys Glu Ala Trp Lys
 165 170 175
 10 Phe Ala Lys Glu Gln Gly Leu Asp Phe Ile Thr Ile Ile Pro Pro Leu
 180 185 190
 Val Val Gly Pro Phe Leu Met Pro Thr Met Pro Pro Ser Leu Ile Thr
 195 200 205
 15 Ala Leu Ser Pro Ile Thr Gly Asn Glu Asp His Tyr Ser Ile Ile Lys
 210 215 220
 20 Gln Gly Gln Phe Val His Leu Asp Asp Leu Cys Leu Ala His Ile Phe
 225 230 235 240
 Leu Phe Glu Glu Pro Glu Val Glu Gly Arg Tyr Ile Cys Ser Ala Cys
 245 250 255
 25 Asp Ala Thr Ile His Asp Ile Ala Lys Leu Ile Asn Gln Lys Tyr Pro
 260 265 270
 30 Glu Tyr Lys Val Pro Thr Lys Phe Lys Asn Ile Pro Asp Gln Leu Glu
 275 280 285
 Leu Val Arg Phe Ser Ser Lys Lys Ile Thr Asp Leu Gly Phe Lys Phe
 290 295 300
 35 Lys Tyr Ser Leu Glu Asp Met Tyr Thr Gly Ala Ile Asp Thr Cys Arg
 305 310 315 320
 Asp Lys Gly Leu Leu Pro Lys Pro Ala Glu Lys Gly Leu Phe Thr Lys
 325 330 335
 40 Pro Gly Glu Thr Pro Val Asn Ala Met His Lys
 340 345
 45 <210> 18
 <211> 342
 <212> PRT
 <213> Arabidopsis thaliana
 50 <400> 18
 Met Asp Gln Thr Leu Thr His Thr Gly Ser Lys Lys Ala Cys Val Ile
 1 5 10 15
 55 Gly Gly Thr Gly Asn Leu Ala Ser Ile Leu Ile Lys His Leu Leu Gln
 20 25 30
 Ser Gly Tyr Lys Val Asn Thr Thr Val Arg Asp Pro Glu Asn Glu Lys
 35 40 45
 60 Lys Ile Ala His Leu Arg Gln Leu Gln Glu Leu Gly Asp Leu Lys Ile

	50		55		60
	Phe Lys Ala Asp Leu Thr Asp Glu Asp Ser Phe Glu Ser Ser Phe Ser				
	65		70		75 80
5	Gly Cys Glu Tyr Ile Phe His Val Ala Thr Pro Ile Asn Phe Lys Ser				
		85		90	95
10	Glu Asp Pro Glu Lys Asp Met Ile Lys Pro Ala Ile Gln Gly Val Ile				
		100		105	110
	Asn Val Leu Lys Ser Cys Leu Lys Ser Lys Ser Val Lys Arg Val Ile				
		115		120	125
15	Tyr Thr Ser Ser Ala Ala Ala Val Ser Ile Asn Asn Leu Ser Gly Thr				
		130		135	140
	Gly Leu Val Met Asn Glu Glu Asn Trp Thr Asp Ile Asp Phe Leu Thr				
		145		150	155
20	Glu Glu Lys Pro Phe Asn Trp Gly Tyr Pro Ile Ser Lys Val Leu Ala				
		165		170	175
25	Glu Lys Lys Ala Trp Glu Phe Ala Glu Glu Asn Lys Ile Asn Leu Val				
		180		185	190
	Thr Val Ile Pro Ala Leu Ile Ala Gly Asn Ser Leu Leu Ser Asp Pro				
		195		200	205
30	Pro Ser Ser Leu Ser Leu Ser Met Ser Phe Ile Thr Gly Lys Glu Met				
		210		215	220
	His Val Thr Gly Leu Lys Glu Met Gln Lys Leu Ser Gly Ser Ile Ser				
		225		230	235
35	Phe Val His Val Asp Asp Leu Ala Arg Ala His Leu Phe Leu Ala Glu				
		245		250	255
40	Lys Glu Thr Ala Ser Gly Arg Tyr Ile Cys Cys Ala Tyr Asn Thr Ser				
		260		265	270
	Val Pro Glu Ile Ala Asp Phe Leu Ile Gln Arg Tyr Pro Lys Tyr Asn				
		275		280	285
45	Val Leu Ser Glu Phe Glu Glu Gly Leu Ser Ile Pro Lys Leu Thr Leu				
		290		295	300
	Ser Ser Gln Lys Leu Ile Asn Glu Gly Phe Arg Phe Glu Tyr Gly Ile				
		305		310	315
50	Asn Glu Met Tyr Asp Gln Met Ile Glu Tyr Phe Glu Ser Lys Gly Leu				
		325		330	335
55	Ile Lys Ala Lys Glu Ser				
		340			
	<210> 19				
	<211> 354				
60	<212> PRT				

<213> Hordeum vulgare

<400> 19

5	Met	Asp	Gly	Asn	Lys	Gly	Pro	Val	Val	Val	Thr	Gly	Ala	Ser	Gly	Phe
	1				5					10					15	
	Val	Gly	Ser	Trp	Leu	Val	Met	Lys	Leu	Leu	Gln	Ala	Gly	Tyr	Thr	Val
				20					25					30		
10	Arg	Ala	Thr	Val	Arg	Asp	Pro	Ala	Asn	Val	Glu	Lys	Thr	Lys	Pro	Leu
			35					40					45			
	Leu	Glu	Leu	Pro	Gly	Ala	Lys	Glu	Arg	Leu	Ser	Ile	Trp	Lys	Ala	Asp
		50					55					60				
15	Leu	Ser	Glu	Asp	Gly	Ser	Phe	Asn	Glu	Ala	Ile	Ala	Gly	Cys	Thr	Gly
		65				70					75					80
	Val	Phe	His	Val	Ala	Thr	Pro	Met	Asp	Phe	Asp	Ser	Gln	Asp	Pro	Glu
20				85						90					95	
	Asn	Glu	Val	Ile	Lys	Pro	Thr	Val	Glu	Gly	Met	Leu	Ser	Ile	Met	Arg
				100					105					110		
25	Ala	Cys	Lys	Glu	Ala	Gly	Thr	Val	Lys	Arg	Ile	Val	Phe	Thr	Ser	Ser
			115					120					125			
	Ala	Gly	Ser	Val	Asn	Ile	Glu	Glu	Arg	Pro	Arg	Pro	Ala	Tyr	Asp	Gln
30				130			135					140				
	Asp	Asn	Trp	Ser	Asp	Ile	Asp	Tyr	Cys	Arg	Arg	Val	Lys	Met	Thr	Gly
		145				150					155					160
	Trp	Met	Tyr	Phe	Val	Ser	Lys	Ala	Leu	Ala	Glu	Lys	Ala	Ala	Met	Glu
35				165						170					175	
	Tyr	Ala	Ser	Glu	Asn	Gly	Leu	Asp	Phe	Ile	Ser	Ile	Ile	Pro	Thr	Leu
			180						185					190		
40	Val	Val	Gly	Pro	Phe	Leu	Ser	Ala	Gly	Met	Pro	Pro	Ser	Leu	Val	Thr
			195					200					205			
	Ala	Leu	Ala	Leu	Ile	Thr	Gly	Asn	Glu	Ala	His	Tyr	Ser	Ile	Leu	Lys
		210				215						220				
45	Gln	Val	Gln	Leu	Val	His	Leu	Asp	Asp	Leu	Cys	Asp	Ala	Met	Thr	Phe
		225				230					235					240
	Leu	Phe	Glu	His	Pro	Glu	Ala	Asn	Gly	Arg	Tyr	Ile	Cys	Ser	Ser	His
50				245						250					255	
	Asp	Ala	Thr	Ile	His	Gly	Leu	Ala	Arg	Met	Leu	Gln	Asp	Arg	Phe	Pro
			260						265					270		
55	Glu	Tyr	Asp	Ile	Pro	Gln	Lys	Phe	Ala	Gly	Val	Asp	Asp	Asn	Leu	Gln
			275					280					285			
	Pro	Ile	His	Phe	Ser	Ser	Lys	Lys	Leu	Leu	Asp	His	Gly	Phe	Ser	Phe
		290					295					300				

60

Arg Tyr Thr Thr Glu Asp Met Phe Asp Ala Ala Ile His Thr Cys Arg
 305 310 315 320
 5 Asp Lys Gly Leu Ile Pro Leu Gly Asp Val Pro Ala Pro Ala Ala Gly
 325 330 335
 Gly Lys Leu Gly Ala Leu Ala Ala Gly Glu Gly Gln Ala Ile Gly Ala
 340 345 350
 10 Glu Thr
 15 <210> 20
 <211> 330
 <212> PRT
 <213> *Oryza sativa*
 20 <400> 20
 Met Val Ile Ser Ser Lys Gly Lys Val Cys Val Thr Gly Ala Ser Gly
 1 5 10 15
 Phe Val Ala Ser Trp Leu Ile Lys Arg Leu Leu Glu Ala Gly Tyr His
 20 25 30
 25 Val Ile Gly Thr Val Arg Asp Pro Ser Asn Arg Glu Lys Val Ser His
 35 40 45
 30 Leu Trp Arg Leu Pro Ser Ala Lys Glu Arg Leu Gln Leu Val Arg Ala
 50 55 60
 Asp Leu Met Glu Glu Gly Ser Phe Asp Asp Ala Val Met Ala Cys Glu
 65 70 75 80
 35 Gly Val Phe His Thr Ala Ser Pro Val Leu Ala Lys Ser Asp Ser Asn
 85 90 95
 Cys Lys Glu Glu Met Leu Val Pro Ala Ile Asn Gly Thr Leu Asn Val
 100 105 110
 40 Leu Lys Ser Cys Lys Lys Asn Pro Phe Leu Lys Arg Val Val Leu Thr
 115 120 125
 45 Ser Ser Ser Ser Thr Val Arg Ile Met Asp Glu Ser Lys His Pro Glu
 130 135 140
 Ile Ser Leu Asp Glu Thr Ile Trp Ser Ser Val Ala Leu Cys Glu Lys
 145 150 155 160
 50 Leu Gln Leu Trp Tyr Ala Leu Ala Lys Ile Ser Ala Glu Lys Ala Ala
 165 170 175
 Trp Glu Phe Ala Lys Glu Asn Asn Ile Asp Leu Val Thr Val Leu Pro
 180 185 190
 55 Ser Phe Val Ile Gly Pro Ser Leu Ser His Glu Leu Ser Val Thr Ala
 195 200 205
 60 Ser Asp Ile Leu Gly Leu Leu Gln Gly Asp Thr Asp Arg Phe Ile Ser
 210 215 220

Tyr Gly Arg Met Gly Tyr Val His Ile Asp Asp Val Ala Ser Cys His
 225 230 235 240
 5 Ile Leu Val Tyr Glu Ala Pro Gln Ala Thr Gly Arg Tyr Leu Cys Asn
 245 250 255
 Ser Val Val Leu Asp Asn Asn Glu Leu Val Ala Leu Leu Ala Lys Gln
 260 265 270
 10 Phe Pro Ile Phe Pro Ile Pro Arg Ser Leu Arg Asn Pro Tyr Glu Lys
 275 280 285
 15 Gln Ser Tyr Glu Leu Asn Thr Ser Lys Ile Gln Gln Leu Gly Phe Lys
 290 295 300
 Phe Lys Gly Val Gln Glu Met Phe Gly Asp Cys Val Glu Ser Leu Lys
 305 310 315 320
 20 Asp Gln Gly His Leu Leu Glu Cys Pro Leu
 325 330
 25 <210> 21
 <211> 341
 <212> PRT
 <213> Fragaria x ananassa
 30 <400> 21
 Met Gly Leu Gly Ala Glu Ser Gly Ser Val Cys Val Thr Gly Ala Ser
 1 5 10 15
 35 Gly Phe Val Gly Ser Trp Leu Val Met Arg Leu Leu Glu His Gly Tyr
 20 25 30
 Thr Val Arg Ala Thr Val Arg Asp Pro Ala Asn Leu Lys Lys Val Arg
 35 40 45
 40 His Leu Leu Glu Leu Pro Gln Ala Ala Thr Arg Leu Thr Leu Trp Lys
 50 55 60
 Ala Asp Leu Asp Val Glu Gly Ser Phe Asp Glu Ala Ile Lys Gly Cys
 65 70 75 80
 45 Thr Gly Val Phe His Val Ala Thr Pro Met Asp Phe Glu Ser Glu Asp
 85 90 95
 50 Pro Glu Asn Glu Val Ile Lys Pro Thr Ile Asn Gly Met Leu Asp Ile
 100 105 110
 Met Lys Ala Cys Leu Lys Ala Lys Thr Val Arg Arg Leu Val Phe Thr
 115 120 125
 55 Ser Ser Ala Gly Ala Val Ala Ile Glu Glu His Pro Lys Glu Val Tyr
 130 135 140
 Ser Glu Asn Asn Trp Ser Asp Val Val Phe Cys Arg Lys Val Lys Met
 145 150 155 160
 60

Thr Gly Trp Met Tyr Phe Val Ser Lys Thr Leu Ala Glu Gln Ala Ala
 165 170 175
 5 Trp Lys Phe Ala Lys Glu Asn Asn Ile Asp Phe Ile Thr Ile Pro
 180 185 190
 Thr Leu Val Ile Gly Pro Phe Leu Ala Pro Ser Met Pro Pro Ser Leu
 195 200 205
 10 Ile Ser Gly Leu Ser Pro Leu Thr Gly Asn Glu Ala His Tyr Gly Ile
 210 215 220
 Ile Lys Gln Cys Gln Tyr Val His Leu Asp Asp Leu Cys Gln Ser His
 225 230 235 240
 15 Ile Phe Leu Tyr Glu His Ala Lys Ala Glu Gly Arg Tyr Ile Cys Ser
 245 250 255
 20 Ser His Asp Ala Thr Ile His Asp Ile Ala Lys Leu Leu Asn Glu Lys
 260 265 270
 Tyr Pro Lys Tyr Asn Val Pro Lys Lys Phe Lys Gly Ile Glu Glu Asn
 275 280 285
 25 Leu Thr Asn Ile His Phe Ser Ser Lys Lys Leu Lys Glu Met Gly Phe
 290 295 300
 Glu Phe Lys His Ser Leu Glu Asp Met Phe Thr Gly Ala Val Asp Ala
 305 310 315 320
 30 Cys Arg Glu Lys Gly Leu Leu Pro Leu Pro Gln Glu Glu Glu Thr Glu
 325 330 335
 35 Lys Arg Arg Ala Gly
 340
 40 <210> 22
 <211> 331
 <212> PRT
 <213> Zea mays
 <400> 22
 45 Met Val Thr Ser Ser Lys Gly Lys Val Cys Val Thr Gly Ala Ser Gly
 1 5 10 15
 Phe Val Ala Ser Trp Leu Ile Lys Arg Leu Leu Glu Ser Gly Tyr His
 20 25 30
 50 Val Val Gly Thr Val Arg Asp Pro Gly Asn His Gln Lys Thr Ala His
 35 40 45
 55 Leu Trp Lys Leu Pro Gly Ala Lys Glu Arg Leu Gln Ile Val Arg Ala
 50 55 60
 Asn Leu Leu Glu Glu Gly Ser Phe Asp Ser Ala Val Met Ala Cys Glu
 65 70 75 80
 60 Gly Val Phe His Thr Ala Ser Pro Val Leu Ala Lys Pro Asp Ser Thr

